# **ANTONELA-CARMEN ARIEŞAN-SIMION**

# INVENTARUL LEXICAL AL LIMBII ROMÂNE CA LIMBĂ STRĂINĂ

**Nivelul A1** 

PRESA UNIVERSITARĂ CLUJEANĂ

#### ANTONELA-CARMEN ARIEŞAN-SIMION

•

### INVENTARUL LEXICAL AL LIMBII ROMÂNE CA LIMBĂ STRĂINĂ NIVELUL A1

## ANTONELA-CARMEN ARIEȘAN-SIMION

# INVENTARUL LEXICAL AL LIMBII ROMÂNE CA LIMBĂ STRĂINĂ NIVELUL A1

#### Referenți științifici:

Prof. univ. dr. emerit G. G. Neamțu Conf. univ. dr. Elena Platon

#### ISBN 978-606-37-1689-8

© 2022 Autoarea volumului. Toate drepturile rezervate. Reproducerea integrală sau parțială a textului, prin orice mijloace, fără acordul autoarei, este interzisă și se pedepsește conform legii.

Universitatea Babeș-Bolyai Presa Universitară Clujeană Director: Codruţa Săcelean Str. Hasdeu nr. 51 400371 Cluj-Napoca, România Tel./fax: (+40)-264-597.401 E-mail: editura@ubbcluj.ro http://www.editura.ubbcluj.ro/

#### LISTĂ DE ABREVIERI

AP – An pregătitor

CECRL – Cadrul european comun de referință pentru limbi: învățare, predare, evaluare

CORLS – Corpus oral de limba română ca limbă străină

E – Erasmus

FA – frecvența absolută

FR – frecvenţa relativă

IL – interlimbă

L1 – prima limbă

L2 – limba a doua

LC – lingvistica bazată pe corpus

LN – limbă naturală

LS – limbă străină

LŢ – limbă-țintă

ML – microlimbă

R – răspândirea

RLM – româna ca limbă maternă

RLNM – româna ca limbă nematernă

RLS – româna ca limbă străină

#### LISTA TABELELOR

Tabelul	1 - Nivelurile comune de referință. Scala globală CECRL	28
Tabelul	2 - Scala lingvistică generală din CECRL	13
Tabelul	3 - Scala din CECRL pentru volumul vocabularului	<b>1</b> 5
Tabelul	4 - Scala din CECRL pentru stăpânirea vocabularului	17
Tabelul	5 – Scala propusă de către Todea pentru vocabularul necesar în citire	77
Tabelul	6 - Numărul de cuvinte aferent fiecărui nivel, pentru limbi diferite 13	30
Tabelul	7 - Tipurile de texte și de teme din corpusul de producții scrise a	le
vorbitor	ilor nonnativi de RLS, Nivelul A114	19
Tabelul	8 – Temele producțiilor scrise	50
Tabelul	9 – Convenții de transcriere în corpusul de producții scrise ale vorbitoril	or
nonnativ	i, Nivelul A1	52
Tabelul	10 – Structura CORLS	53
Tabelul	11 - Grila de evaluare pentru producerea scrisă	59
Tabelul	12 - Primele 10 cuvinte din lista de frecvențe, versiunea 1 16	51
Tabelul	13 – Lista numelor proprii care se întâlnesc în corpusul de producții scri	se
ale vorb	itorilor nonnativi de RLS	72
Tabelul	14-Lista acronimelor și a abrevierilor care apar în corpusurile vorbitoril	or
nonnativ	vi de RLS	72
Tabelul	15 – Lista numerelor care apar în corpusurile vorbitorilor nonnativi de RI	_S
		73
Tabelul	16 – Elemente ce aparțin discursului oral, excluse din analiză	75
Tabelul	17 – Transferuri directe dintr-o limbă străină	76
	18 – Transferuri directe dintr-o limbă străină	
Tabelul	19 – Lista sferelor lexicale	39
Tabelul	20 – Elementele lexicale propuse pe baza criteriului necesității 19	€0
Tabelul	21 – Excluderea unor cuvinte pe baza principiului melcului de mare 19	€1
Tabelul	22 – Primele 11 cuvinte din lista finală de frecvențe	€
Tabelul	23 - Inventarul lexical al limbii române ca L2, nivelul A1	€7

#### LISTA FIGURILOR

Figura	1 - Structura schemei descriptive a CECRL	25
Figura	2 - Nivelurile comune de referință	27
Figura	3 - Noua schemă descriptivă a competențelor lingvistice de comunicare	39
Figura	4 - L1 în cazul studenților din an pregătitor, în perioada 2013-2016 (stá	ànga),
respect	iv, 2013-2018 (dreapta)	100
Figura	5 - Scenariu mental pentru <i>a lua masa în oraș</i>	112
Figura	6 – Distribuția L1	145
Figura	7 - Distribuția pe grupe de vârstă	146
Figura	8 - Distribuția în funcție de genuri	147
Figura	9 - FA în corpusurile vorbitorilor nonnativi de română ca L2	161
Figura	10 – Legea lui Zipf, Lista de frecvență v2	194
Figura	11 – Legea lui Zipf – primele 400 de cuvinte	195
Figura	12 – Legea lui Zipf – primele 500 de cuvinte	195
Figura	13 – Legea lui Zipf – primele 1000 de cuvinte	196

#### **CUPRINS**

Cuvânt-înainte	12
SECȚIUNEA I - CADRU TEORETIC	17
Capitolul I – Rolul documentelor lingvistice europene în procesul de aliniere a limromâne ca limbă străină la standardele internaționale	
1.1. Cadrul european comun de referință pentru limbi: învățare, predare, evalu	
1.2. Cadrul european comun de referință pentru limbi: învățare, predare, evalua Companion Volume	
1.3. Materiale elaborate pornind de la CECRL	36
1.4. Competența lexicală	.39
1.5. Concluzii	.47
Capitolul II – Observații asupra stadiului actual al cercetării	49
2.1. Situația altor limbi de circulație internațională	.50
2.2. Situația actuală a cercetării pe teren românesc	.57
Capitolul III – Despre necesitatea și relevanța inventarelor lexicale	. 72
3.1. Despre necesitatea ierarhizării lexicului RLS	.74
3.2. Perspectiva CECRL asupra lexicului ierarhizat pe niveluri	.79
3.3. Inventar lexical vs dicționar	.81
3.4. Utilitatea inventarelor lexicale în procesele de predare/învățare/evaluare	.82
3.5. Concluzii	.86
Capitolul IV – Corpusul lingvistic, în general	. 87
4.1. Corpusul lingvistic. Clasificări	.90
4.2. Avantajele unui corpus de producții scrise ale vorbitorilor nonnativi	.94
4.3. Lingvistica bazată pe corpus. Aplicabilitate interdisciplinară	105
4.4. Concluzii	107
Capitolul V – Vocabularul limbii române. Cadru general 1	108

5.1. Vocabularul limbii române. Considerații generale
5.2. Perspective în abordarea cunoașterii lexicale în L2
5.3. Principii de urmat în elaborarea listelor de cuvinte ierarhizate116
5.4. Criterii
5.5. Delimitarea numărului de cuvinte pe fiecare nivel de competență lingvistică
5.6. Concluzii
Capitolul VI – Rolul statisticii în lingvistica bazată pe corpus (LC)
6.1. Statistica și LC
6.2. Corpusul de producții scrise ale vorbitorilor nonnativi de limba română ca limbă străină. Nivelul A1140
6.3. Descrierea corpusului de producții scrise ale vorbitorilor nonnativi, nivelul A1143
6.4. CORLS - Corpus oral de limba română ca limbă străină152
6.5. Corpusul de cerințe
6.6. Concluzii
SECȚIUNEA A II-A – REZULTATELE CERCETĂRII
Capitolul VII – Rezultatele cercetării
7.1. Metodologie
7.2. Analiza cantitativă160
7.3. Substantivele proprii
7.4. Acronime și abrevieri172
7.5. Numerele
7.6. Excluderi din lista de frecvențe
7.7. Analiza calitativă178
7.8. Criteriul întregirii sferelor lexicale
7.9. Subcriteriul necesității190
7.10. Legea lui Zipf aplicată listei de frecvențe din corpusul de producții scrise ale vorbitorilor nonnativi de RLS

7.11. Concluzii	196
Concluzii	203
Bibliografie	208
Anexe	221
Anexa 1 - Lista substantivelor proprii, a abrevierilor și a acronimelor.	
Anexa 2 - Lista formulelor memorate, nivelul A1	227
Anexa 3 - Corpus scris de limba română ca limbă străină Nivelul A1	230
Anexa 4 – Lista de frecvență, versiunea finală	492

#### **CUVÂNT-ÎNAINTE**

Dacă, în spațiul lingvistic internațional, atenția cercetătorilor se îndreaptă din ce în ce mai mult asupra realizării unor descrieri cât mai complete ale limbilor, în literatura română de specialitate întâlnim puține materiale destinate acestei problematici, iar dintre acestea și mai puține sunt acelea care își bazează cercetarea pe un corpus de producții ale vorbitorilor nonnativi de limba română.

Prin această lucrare, tratăm o temă actuală, însă puţin abordată de către specialiștii români: descrierea, din punct de vedere lexical, a limbii române ca limbă a doua (L2), la nivelul A1, prin realizarea unui inventar lexical aferent nivelului A1 din *Cadrul european comun de referință pentru limbi* (CECRL), elaborat pe baza unui corpus al vorbitorilor nonnativi (*learner corpus*). Totuși, necesitatea acestei teme a fost subliniată, în mai multe rânduri, de către diferiți specialiști români, precum Elena Platon, Gabriela Biriș, Diana Burlacu, Gabriela Stoica, Adriana Todea sau Carmen Mîrzea Vasile, după cum am arătat în corpul lucrării.

Întrucât unele dintre obiectivele principale ale cercetării românești în domeniul RLS le constituie sincronizarea cu tendințele internaționale, în scopul alinierii la politicile lingvistice europene, dar și acela de a beneficia de experiența acumulată în situația altor limbi, am considerat oportună propunerea unei astfel de teme, cu relevanță majoră, în opinia noastră, pentru cercetările viitoare desfășurate în acest domeniu. Așadar, noutatea temei propuse o reprezintă tocmai realizarea unei cercetări de acest fel pentru limba română ca LS/L2. Din punctul nostru de vedere, un astfel de instrument lingvistic și didactic oferă o multitudine de avantaje atât pentru specialiști, cât și pentru toate persoanele interesate de studiul limbii române. În continuare, prezentăm principalele beneficii pe care considerăm că le aduce inventarul lexical al limbii române ca LS/L2, pentru nivelul A1:

- 1. Reprezintă, în plan științific, un material relevant pentru descrierea și cercetarea diferitelor aspecte ale limbii române, ca L1, L2 sau LS, având în vedere că este elaborat pe baza corpusului vorbitorilor nonnativi.
- Constituie un instrument umportant în activitățile de planificare a cursurilor, dar și de elaborare a materialelor necesare predării, învățării și evaluării limbii române ca LS/L2.
- 3. Reprezintă un reper esențial în elaborarea cărților adaptate pe nivel, materiale de mare succes în cazul altor limbi străine, dar care, deocamdată, în cazul limbii române sunt disponibile într-un număr redus.

- 4. Asigură un grad ridicat de unitate în ceea ce privește procesele de predare și evaluare, preîntâmpinând, astfel, situațiile în care, la evaluare, de exemplu, pot apărea diferențe din cauza utilizării unor materiale de pregătire foarte diferite. Pe de altă parte, și în ceea ce privește evaluarea propriu-zisă, aduce un mare avantaj, stabilind aceleași cerințe, pe plan lexical, indiferent de centrul unde are loc testarea.
- 5. Susține și direcționează procesul de învățare și autoevaluare, în cazul autodidacților.
- 6. Constituie un material-sursă, pe baza căruia se pot elabora diferite activități sau materiale destinate dezvoltării vocabularului, de exemplu, cartonașe de tip *flashcards*, jocuri de memorie sau chiar jocuri și aplicații pentru telefon sau pentru calculator.

După cum se poate observa din expunerea beneficiilor aduse de *Inventarul lexical al limbii române*, nu încurajăm învățarea vocabularului sub formă de liste de cuvinte. În cazul analizei noastre, s-a dovedit, într-adevăr, că aceste elemente lexicale sunt mai ușor de însușit (cu grupul-țintă descris și în condițiile descrise), dar acest lucru nu presupune că acestea vor fi singurele predate și învățate, ci înseamnă că acest inventar dorește să ofere un reper, din perspectiva lexicului, în special în ceea ce privește evaluarea, la nivelul A1.

Un alt beneficiu major pentru domeniul RLS/al achiziției L2 este reprezentat de faptul că, odată cu inventarul lexical, cercetătorii vor dispune de un alt material, la fel de util și ofertant pentru cercetările viitoare. Ne referim, desigur, la corpusul de producții scrise ale vorbitorilor nonnativi, compus din textele redactate la examenul organizat la finalul nivelului A1 de către studenții străini din cadrul programelor universitare de An pregătitor și Erasmus de la Universitatea Babeș-Bolyai din Cluj-Napoca. Beneficiile aduse de acest corpus, alături de cel oral (CORLS), elaborat de către Lavinia-Iunia Vasiu, nu se limitează doar la domeniul RLS, ci vor putea fi observate și în cel al RLM.

Din punctul de vedere al conținutului, lucrarea noastră este formată din două secțiuni, diferite ca întindere: o primă parte, dedicată *Cadrului general* și aspectelor de ordin teoretic, și a doua parte, practică, în cadrul căreia sunt prezentate rezultatele analizei noastre. Partea teoretică este segmentată în șase capitole, după cum urmează:

Capitolul 1, Rolul documentelor lingvistice europene în procesul de aliniere a limbii române ca limbă străină la standardele internaționale, în care propunem o prezentare a principalelor documente lingvistice europene, dar și a principiilor și a conceptelor pe care le promovează, evidențiindu-se avantajele și neajunsurile pe care acestea le comportă.

- Am încheiat acest capitol cu discutarea competenței lexicale, așa cum este descrisă în *CECRL* și în *Companion Volume*.
- ❖ Capitolul 2, Observații asupra stadiului actual al cercetării, este destinat descrierii stării actuale a domeniului, atât pe plan intern, cât și internațional. Astfel, am prezentat succint lucrările de referință pentru tema noastră, arătând beneficiile pe care le-au adus în momentul apariției acestora, precum și limitele acestor materiale, din punctul nostru de vedere.
- ❖ Capitolul 3, Despre necesitatea și relevanța inventarelor lexicale, continuă demersul nostru și explică rațiunea din spatele propunerii acestei teme pentru domeniul limbii române ca LS/L2, detaliind motivele pentru care un inventar al limbii române ca LS/L2 este deosebit de util. Tot în cadrul acestui capitol se arată care sunt principalele diferente dintre un inventar si un dictionar.
- ❖ Capitolul 4, *Corpusul lingvistic*, *în general*, tratează, după cum se poate observa din titlu, problema corpusurilor lingvistice, evidențiind principalele astfel de materiale electronice existente pentru limba română, dar și clasificările operate în cazul corpusurilor. Tot aici discutăm despre analiza lingvistică bazată pe corpus și despre aplicabilitatea sa interdisciplinară.
- ❖ Capitolul 5, *Vocabularul limbii române*. *Cadru general*, începe cu o prezentare a vocabularului limbii române și se continuă cu accepțiunile privind învățarea lexicală, atât în L1, cât și în L2. Nu în ultimul rând, în acest capitol arătăm care sunt liniile directoare în elaborarea listelor de cuvinte sau a inventarelor lexicale, așa cum apar în modelele internaționale, de sorginte anglofonă.
- ❖ Capitolul 6, Rolul statisticii în lingvistica bazată pe corpus (LC), face trecerea înspre partea aplicată a lucrării și prezintă principalele metode și concepte statistice în LC. De asemenea, în acest capitol, descriem și corpusul de producții scrise ale vorbitorilor nonnativi de limba română, nivelul A1. Propunem, de asemenea, o scurtă prezentare și a celorlalte corpusuri utilizate pentru a avea o privire de ansamblu asupra microlimbii (ML= limba simplificată față de limba naturală, folosită în general, vehiculată de către vorbitorul nativ în interacțiunea cu un nonnativ, în scopul asigurării succesului comunicării) de nivel A1 și pentru a obține niște rezultate cât mai relevante.

În secțiunea a 2-a, *Rezultatele cercetării*, prezentăm metodologia și etapele parcurse până la obținerea inventarului lexical pentru nivelul A1. În această secțiune, am evidențiat ceea ce poate face și ceea ce înțelege studentul, din punct de vedere lexical, și, implicit, ce elemente lexicale ar presupune microlimba și interlimba la nivelul A1. Aici, am izolat mai multe liste de cuvinte, cu rol de instrumente în procesele de proiectare a materialelor/a cursurilor. Pentru a face cât mai accesibil inventarul pentru utilizator, dar mai ales pentru rolul său de instrument didactic, am propus suplimentarea listei principale cu: tabelul numelor proprii, a acronimelor și a abrevierilor, precum și cu o listă de formule memorate. În plus, deși au fost incluse în inventar, am realizat și separat o listă a câtorva sfere lexicale întâlnite în corpusurile vorbitorilor nonnativi. Aceasta reprezintă doar un model, și nu are pretenția de a avea un caracter exhaustiv, de a acoperi toate sferele lexicale abordate la nivelul A1.

Cele două secțiuni sunt urmate de *Concluziile* generale și de *Anexe*, în care se găsesc propuneri de liste, utilizate în scop didactic, și corpusul de producții scrise ale vorbitorilor nonnativi de nivelul A1, în care, primele pagini sunt reprezentate de câteva metadate.

Întrucât o cercetare nu poate fi considerată completă până ce nu trece și printr-o etapă critică, am încercat să facem un exercițiu de autocritică, identificând eventualele reproșuri ce ar putea fi aduse lucrării noastre. Un prim aspect ar putea fi legat de dimensiunea corpusului folosit. Desi recunoaștem că 56.000 de elemente nu reprezintă o dimensiune mare, amintim opiniile cercetătorilor internaționali care evidențiau că dimensiunea corpusului trebuie să răspundă cerințelor cercetării, iar, în cazul nostru, considerăm că o face. Pe de altă parte, având în vedere caracterul divers al subiecților noștri, subliniem, încă o dată, dificultatea cu care s-au colectat datele, totul depinzând de acceptul studenților privind folosirea producțiilor, dar și imposibilitatea includerii unora dintre acestea din cauza caracterului neinteligibil. În plus, deoarece am urmărit descrierea vocabularului limbii române ca limbă străină, în general, am inclus, în analiza noastră, și corpusul de producții orale, ceea ce a dus la mai mult decât dublarea dimensiunii corpusului. Odată cu acest aspect, răspundem unui alt posibil reproș, legat de diversitatea redusă a temelor de la scriere. Considerăm că această problemă a fost soluționată odată cu subiectele din corpusul oral, asigurându-se un grad bun de acoperire a temelor care caracterizează nivelul A1, conform CECRL, cu privire la detalii personale și nevoi concrete imediate.

Un alt aspect ce ni s-ar putea imputa este legat de modalitatea de colectare a producțiilor, și anume selectarea exclusivă a celor din cadrul examenelor, situație care, după unii, ar presupune automat lipsa autenticității. Chiar dacă suntem de acord

cu faptul că situațiile de examen pot reprezenta o sursă suplimentară de stres, care ar putea influența producțiile, totuși acestea reprezintă și ele mostre de limbă naturală, după cum evidențiau Ellis și Barkhuizen și după cum am arătat în corpul lucrării. Pe de altă parte, prin colectarea textelor la final de A1, ne-am asigurat de o bună reprezentare a vocabularului, însă, chiar și în acest moment, elementele lexicale sunt scrise, de foarte multe ori, cu mari erori, după cum se poate observa în corpus. Considerăm că o colectare și mai precoce nu ar fi fost relevantă, pe de-o parte, deoarece, probabil, ar fi conținut cuvinte sau enunțuri scurte și scrise greșit, iar, pe de altă parte, nu credem că ar fi putut demonstra însușirea elementelor lexicale, având în vedere că vorbim despre o perioadă destul de scurtă de învătare.

Un alt reproş ar putea fi legat de trasăturile grupului nostru țintă, în majoritate, vorbitor de limbă arabă. Într-adevăr, avem un grup de subiecți printre care domină limba arabă, însă, prin includerea studenților din programul universitar Erasmus, considerăm că am asigurat o bună reprezentare și a vorbitorilor cu alte L1, precum germana, franceza și italiana. Mai mult, avem convingerea că, așa cum este descris grupul nostru țintă, acesta este reprezentativ, în general, pentru studenții străini de la Universitatea Babeş-Bolyai din Cluj-Napoca.

Prin urmare, considerăm că, prin lucrarea noastră, aducem un beneficiu domeniului, în special prin cele două instrumente propuse, inventarul lexical și corpusul de producții scrise ale nonnativilor, pentru nivelul A1.

# SECȚIUNEA I - CADRU TEORETIC

# CAPITOLUL I – ROLUL DOCUMENTELOR LINGVISTICE EUROPENE ÎN PROCESUL DE ALINIERE A LIMBII ROMÂNE CA LIMBĂ STRĂINĂ LA STANDARDELE INTERNAȚIONALE

#### 1.1. Cadrul european comun de referință pentru limbi: învățare, predare, evaluare

Cadrul european comun de referință pentru limbi: învățare, predare, evaluare (denumit în continuare CECRL)<sup>1</sup> constituie, "probabil, cel mai puternic document european care are drept scop traducerea și învățarea limbilor" [tr. n.]<sup>2</sup>. Acesta este și motivul pentru care am ales să începem demersul nostru cu o scurtă prezentare a CECRL, precum și a celorlalte instrumente lingvistice europene care au drept punct de plecare acest document de referință.

Dorința unei minime standardizări a procesului de predare-evaluare în cazul unei limbi străine, dar și necesitatea de a avea un document de referință pentru toate limbile și pentru toți beneficiarii au stat la baza apariției CECRL. Meritul pentru această inițiativă îl are Consiliul Europei (Divizia de Politici Lingvistice), care, în perioada 1989-1996, a întrunit echipa responsabilă pentru elaborarea primelor versiuni ale acestui document lingvistic european de cea mai mare importanță la nivelul organizării politicilor lingvistice, urmând ca, în anul 2001, după colectarea feedbackului și modificarea variantelor inițiale în funcție de acesta, să apară, în limbile engleză și franceză, versiunea finală a CECRL³. Momentul apariției propriu-zise a fost unul extrem de important, pe mai multe planuri, întrucât a fost nevoie de un cadru comun care să cuprindă toate politicile lingvistice din institutiile de educatie

<sup>&</sup>lt;sup>1</sup> Consiliul Europei, Diviziunea Politici Lingvistice, Strasbourg, *Cadrul European Comun de referință pentru Limbi: Învățare, Predare, Evaluare*, traducere coordonată și revizuită de Gheorghe Moldovanu, Chișinău, 2003. Din păcate, până la acest moment, nu s-a realizat o altă traducere a CECRL, pe teritoriu românesc, iar cea existentă, elaborată de către colectivul din Republica Moldova, nu constituie întotdeauna o reprezentare naturală a limbii române și nici o traducere fidelă a versiunii originale. De aceea am optat pentru utilizarea unei versiuni originale, cea în limba engleză: Council of Europe, Modern Languages Division, Strasbourg, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge, Cambridge University Press, 2001, disponibil în format electronic aici: https://rm.coe.int/1680459f97.

<sup>&</sup>lt;sup>2</sup> A. Müller-Hartmann, M. Schocker-von Ditfurth, *Introduction to English language Teaching*, Stuttgart, Klett Lerntraining, 2010, p. 53.

<sup>&</sup>lt;sup>3</sup> Pentru mai multe detalii privind etapele elaborării CECRL, vezi primul capitol din: Lavinia-Iunia Vasiu, *Achiziția limbii române ca L2. Interlimba la nivelul A1*, Cluj-Napoca, Editura Presa Universitară Clujeană, 2020, pp. 23-49.

din Europa. În special, s-a dorit ca, astfel, să se poată recunoaște, la nivel internațional, calificările lingvistice, obținute în urma absolvirii unor cursuri de limbi sau ca rezultat al susținerii examenelor de competență lingvistică. Așadar, CECRL este văzut ca un punct de plecare, ca un element comun care stă la baza fundamentării unei multitudini de alte documente și materiale, care descriu și clasifică, sub principiile oferite de CECRL, limbile. După Vasiu, acest material:

"se dorește a fi înainte de toate un instrument de planificare și organizare, care propune o *lingua franca* pentru descrierea obiectivelor, a metodelor de predare și a evaluării în domeniul limbilor străine predate și învățate în contexte diverse, facilitând dezvoltarea syllabusurilor, a manualelor, a testelor și a programelor de perfecționare pentru profesori și stimulând la reflecție și discuții"<sup>4</sup>.

Şi North, unul dintre autorii CECRL, susține că "progresul fundamental adus de CECRL vine din faptul că acum oamenii se înțeleg unul pe celălalt mult mai bine, indiferent de cultura pedagogică sau de alte aspecte (testare, perfecționare etc.) și sunt capabili să discute mai eficient și cu rezultate mai bune aceste aspecte" [tr. n.]<sup>5</sup>. Pe de altă parte, în plan didactic, atât în context european, cât și noneuropean<sup>6</sup>, pe lângă rolul său de document lingvistic de referință, acesta a marcat și mutarea accentului de pe *metoda structurală*, o învățare lineară, bazată pe însușirea structurilor gramaticale și a noțiunilor predefinite, pe cea *comunicativ-funcțională*, fundamentată pe un syllabus bazat pe analize de nevoi și orientat înspre propunerea unor sarcini autentice, care pot fi realizate cu ajutorul unor structuri gramaticale, elemente de vocabular și funcții determinate prin procedee empirice: "adoptarea funcțiilor comunicative de către Consiliul Europei a condus meseria de profesor înspre schimbarea ce a avut loc în perioada anilor 1970-1980, când dascălii au început să fie mai preocupați de modalitățile în care poate fi folosită limba în situațiile reale de comunicare" [tr. n.]<sup>7</sup>.

Făcând referire la metafora Turnului Babel, folosită cu alte ocazii de North pentru a justifica rolul CECRL, Sonea afirmă că "ideea din spatele CECRL nu a fost

<sup>5</sup> Brian North, "The CEFR illustrative descriptor scales", in *Modern Language Journal*, no. 91, 2007, p. 659.

<sup>&</sup>lt;sup>4</sup> Lavinia-Iunia Vasiu, op. cit., 2020, pp. 23-24.

<sup>&</sup>lt;sup>6</sup> Conform paginii oficiale, CECRL a fost tradus în 40 de limbi, majoritatea europene, dar și noneuropene, de exemplu: chineză, japoneză, coreeană, arabă sau ebraică.

<sup>&</sup>lt;sup>7</sup> Anthony Green, "The theoretical foundations for functions in the Council of Europe modern languages projects and the Common European Framework of Reference for languages", in Michael Milanovic, Nick Saville (eds.), *Language Functions Revisited*, Cambridge, Cambridge University Press, 2012, p. 7.

aceea de a se întoarce la o perioadă pre-Babel, când toți oamenii vorbeau aceeași limbă, ci de a câștiga o formă de glossolalie, capacitatea de a vorbi limbi diferite, și totuși de a se înțelege reciproc" [tr. n.]<sup>8</sup>. În contextul discuției noastre, considerăm extrem de relevantă afirmația lui Anthony Green privind apariția CECRL: "[acesta] nu a apărut peste noapte, ci a fost rezultatul a peste 30 de ani de proiecte ale Consiliului Europei, dar și a mai îndelungatelor tradiții de analiză lingvistică și de pedagogie a limbilor, toate având drept preocupare principală limba ca mijloc de comunicare" [tr. n.]<sup>9</sup>. În acest parcurs, dinspre metoda structurală înspre cea comunicativă, un punct-cheie a fost reprezentat de apariția *The Threshold Level*<sup>10</sup>—moment crucial, de altfel, și în lingvistica aplicată —, care a descris pentru prima dată ce presupune primul stadiu în procesul de însușire a limbii a doua (de acum înainte, L2), când cunoașterea limbii e la un nivel într-atât de dezvoltat încât vorbitorul poate face față situațiilor de zi cu zi<sup>11</sup>. Constituind un material esențial în achiziția L2, *The Treshold Level* a fost tradus (și adaptat) și în limba română în anul 2001, sub forma *Nivelului Prag*<sup>12</sup>.

Am ales să folosim conceptul de L2, bazându-ne pe diferențierea pe care o fac Webb și Nation între limba maternă, limba străină și limba a doua. Aceștia susțin că felul în care învățăm limba maternă diferă în funcție de maniera în care ne însușim cuvintele dintr-o altă limbă. În limba maternă suntem expuși în permanență atât la conținuturi scrise, cât și orale, în vreme ce, în contextul limbii a doua (adică învățarea L2 în mediu vorbitor, într-o țară în care această limbă e în general folosită), studenții pot primi un input 13 bogat în L2, însă totul depinde de motivația indivizilor de a comunica în limba-țintă (LŢ). Pe de altă parte, într-un context în care se învață o limbă ca limbă străină (adică într-o țară în care nu este în general vorbită), studenții primesc un input mult mai redus și, de cele mai multe ori, în cadru instituționalizat, la cursuri 14. Având în vedere această diferențiere, considerăm că descrierea L2 caracterizează cel mai bine situația limbii române predate și învățate la

<sup>&</sup>lt;sup>8</sup> Sonea, Ioana, "The impact of the CEFR on Teaching and Assessing Romanian as a Second/Foreign Language. Practice and Research", in Lucia-Mihaela Grosu-Răduleascu (ed.), *Foreign Language Teaching in Romanian Higher Education. Teaching Methods, Learning Outcomes*, Cham, Springer, col. "Multilingual Education", no. 27, 2018, p. 85.

<sup>&</sup>lt;sup>9</sup> Anthony Green, op. cit., 2012, p. 7.

<sup>&</sup>lt;sup>10</sup> J. A. van Ek, *The Threshold Level*, Council of Europe, 1975.

<sup>&</sup>lt;sup>11</sup> J. A. van Ek, J.L.M. Trim, *Threshold 1990*, Cambridge, Cambridge University Press, 1991, p. 2.

<sup>&</sup>lt;sup>12</sup> Victoria Moldovan, Liana Pop, Lucia Uricaru, *Nivel prag pentru învățarea limbii române ca limbă străină*, Strasbourg, Consiliul de Cooperări Culturale, 2001.

<sup>&</sup>lt;sup>13</sup> Prin *input* înțelegem limba pe care o receptează vorbitorul nonnativ, iar prin *output*, producția vorbitorului nonnativ. Vezi, în acest sens, Lavinia Iunia Vasiu, *op. cit.*, 2020, *Capitolul II – Achiziția limbii a doua (AL2)*.

<sup>&</sup>lt;sup>14</sup> Stuart Webb, Paul Nation, *How Vocabulary is Learned*, London, Oxford University Press, 2017, p. 50.

Departamentul de limbă, cultură și civilizație românească, de la Facultatea de Litere a Universității Babeș-Bolyai din Cluj-Napoca, întrucât studenții străini, locuind pe perioada studiilor în România, sunt în permanență expuși la LŢ.

Rolul extrem de important jucat de CECRL în standardizarea proceselor de predare, învățare și evaluare ale unei limbi este evidențiat și în volumul *Using the CEFR Principles of Good Practice*<sup>15</sup>. În acest scop, se aduce în discuție chiar un prim pasaj din *CECRL* care arată că:

"acest document a fost realizat cu intenția de a fi un material mereu adaptabil, nu un standard internațional [ceea ce e considerat tot mai des] sau o autorizație acordată de o instituție superioară. Ar trebui privit ca un ghid general, nu ca un instrument care stabilește rețete și, [mai mult], nu oferă răspunsuri gata pregătite sau o metodă unică spre a fi aplicate" [tr.n.]<sup>16</sup>.

Așadar, este important să cunoaștem ideile generale lansate și susținute de autorii CECRL, însă e nevoie să adaptăm aceste principii nevoilor limbii noastre, realizând niște linii de demarcație, care pot fi depășite, în funcție de caz. Mai departe, această idee este susținută printr-o alta, la fel de elocventă: "Cadrul se dorește a fi nu numai inteligibil, transparent și coerent, ci și deschis, dinamic și nondogmatic." [tr.n.]<sup>17</sup>. De exemplu, luând în discuție tema cercetării noastre, dacă privim inventarele lexicale pe niveluri din perspectiva predării și a învățării (în mediu vorbitor sau nu), granițele dintre niveluri ar putea să fie ușor depășite, însă, dacă ne raportăm la evaluare, nu mai este la fel de indicat să încălcăm aceste limite, întrucât variabilele implicate în acest proces sunt mult mai complexe, iar, pentru ca un test să fie standardizat și să-și atingă obiectivele, trebuie să fie astfel realizat încât să asigure candidaților, din diferite centre universitare sau celor care au învățat româna în varii contexte, șanse egale.

După cum s-a putut observa din cele prezentate până aici, viziunea adoptată de CECRL este una "axată pe acțiune"<sup>18</sup>, cel care învață o limbă străină este perceput ca "actor social (membru al societății), care are de îndeplinit diferite sarcini (nu neapărat în întregime de natură lingvistică), într-o serie de circumstanțe date, într-un mediu specific și într-un domeniu particular de activitate" [tr.n.]<sup>19</sup>.

<sup>&</sup>lt;sup>15</sup> Using the CEFR Principles of Good Practice, October 2011, Cambridge ESOL, Cambridge.

<sup>&</sup>lt;sup>16</sup> *Ibidem*, p. 2.

<sup>&</sup>lt;sup>17</sup> *Ibidem*, p. 3.

<sup>&</sup>lt;sup>18</sup> CECRL, p. 9.

<sup>&</sup>lt;sup>19</sup> Ibidem.

Asadar, prin abordarea asumată, în general, precum și prin celelalte teme de discuție propuse, în particular (a se vedea, de exemplu, acceptiunea asupra evaluării), se evidentiază destul de clar abordarea cu predilectie comunicativă a CECRL<sup>20</sup>.

Fundamentală pentru domeniul predării-evaluării limbilor ni se pare și remarca lui David Little despre CECRL, care spune că, grație abordării asumate, acesta "ne pune la dispoziție o definiție a performanței lingvistice în L2, în funcție de limba folosită" [tr.n.]<sup>21</sup>, după cum urmează:

"utilizarea unei limbi, inclusiv învătarea ei, cuprinde un sir de actiuni îndeplinite de către persoane care, în calitate de indivizi și de actori sociali, își dezvoltă un ansamblu de competente generale, dar, mai ales, o competentă de comunicare bazată pe exploatarea mijloacelor lingvistice. Ei utilizează competentele de care dispun în contexte si condiții variate, conformându-se diferitelor constrângeri în vederea realizării activităților comunicative, care permit receptarea și producerea textelor pe anumite teme asupra unor domenii specifice, aplicând strategiile corespunzătoare cele mai adecvate, pentru îndeplinirea sarcinilor date. Controlul acestei activități de către interlocutori conduce la consolidarea sau modificarea competențelor" [tr.n.]<sup>22</sup>.

În Figura 1 se poate observa structura schemei descriptive din CECRL, care are ca nucleu performanta lingvistică, asa cum a fost schitată de E. Piccardo et alli. 23

<sup>22</sup> CECRL, p. 9.

<sup>&</sup>lt;sup>20</sup> David Little, "Common European Framework of Reference for languages; content, purpose, origin, reception and impact", în Language Teaching, no. 39, 2006, p. 169.

<sup>&</sup>lt;sup>21</sup> David Little, The Common European Framework of Reference for Languages, the European Language Portfolio, and language learning in higher education, p. 3, disponibil în format electronic pe: http://languagesinitiative.ie/images/CEFR ELP.pdf.

<sup>&</sup>lt;sup>23</sup> Preluată și tradusă din E. Piccardo et alli, *Pathways Through Assessment, Learning and Teaching in* the CEFR, Graz, Austria, European Centre for Modern Languages, 2011, p. 55.

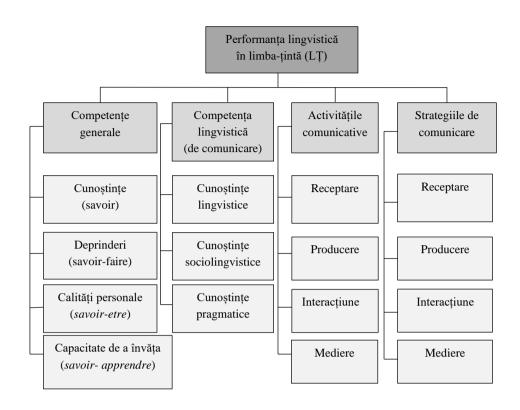


Figura 1 - Structura schemei descriptive a CECRL

După cum s-a putut remarca din sublinierile noastre, odată cu CECRL apar o serie de concepte fundamentale pentru domeniul limbilor străine. În cele ce urmează, vom încerca să le prezentăm într-un mod cât mai succint cu putință. Ne oprim, în primul rând, la termenul *competență*<sup>24</sup>, care e văzut, aici, ca "o sumă a cunoștințelor, abilităților și a caracteristicilor care îi permit unei persoane să întreprindă diferite acțiuni" [tr.n.]<sup>25</sup>. Această definiție pare apropiată de aceea dată de către Dell Hymes conceptului de *competență comunicativă*: "ansamblul cunostințelor lingvistice, interacționale și culturale, internalizate de către vorbitorul

<sup>&</sup>lt;sup>24</sup> Atunci când analizăm schema descriptivă din CECRL, e simplu de observat că perspectiva asupra conceptului de *competență* își are originile în mai multe domenii, nu doar în lingvistica aplicată, acordându-se atenție și aspectelor de ordin sociolingvistic sau domeniului psihologiei aplicate (vezi, pentru mai multe detalii: Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*, Language Policy, 2018, p. 130).

<sup>&</sup>lt;sup>25</sup> CECRL, p. 9.

nativ al unei limbi, care îi permite să se manifeste adecvat în contexte comunicative specifice"<sup>26</sup>. Dacă vorbim despre "unele acțiuni care nu sunt neapărat specifice comunicării, ci se referă mai degrabă la toate tipurile de actiuni (însă poate face referire și la aspectul lingvistic)" [tr.n.]<sup>27</sup>, avem de-a face cu conceptul de *competență* generală. Spre deosebire de această competență generală, competența lingvistică (de comunicare) este aceea care "îi permite unei persoane să acționeze, folosind în special mijloace lingvistice" [tr.n.]<sup>28</sup>. După cum este lesne de dedus, acest din urmă concept ne interesează în mod special în demersul nostru. În CECRL (vezi Figura 1), performanța e văzută drept "abilitatea de a lua parte la activități de comunicare – ilustrată prin descriptori de tipul pot să (can do), în timp ce se folosesc atât competențele generale, cât și cele lingvistice, activându-se strategiile comunicative necesare" [tr.n.]<sup>29</sup>. Acestea din urmă sunt văzute, asadar, drept "liantul dintre activitățile comunicative și competența lingvistică de comunicare" [tr.n.]<sup>30</sup>. După CECRL, prin activitățile comunicative (language activities) înțelegem "implicarea competentei lingvistice de comunicare într-un domeniu specific în scopul procesării (receptării sau producerii) unuia sau a mai multor texte în vederea îndeplinirii unei sarcini" [tr.n.]<sup>31</sup>. Iar, prin *context*, înțelegem "multitudinea de evenimente, precum și de factori situaționali (atât de ordin extern, cât și proprii utilizatorului) în care apar actele de vorbire" [tr.n.]<sup>32</sup>. Pe lângă această acceptiune, vom mai folosi același termen și în sintagma context lingvistic, cu sensul de "parte a unui enunț (sau părți ale unui enunt) care precedă și/sau urmează unitatea lingvistică supusă analizei și care îi condiționează prezența, forma sau funcția"33.

Dacă ar fi să privim CECRL drept un sistem elaborat pe două axe, una orizontală și alta verticală, conceptele prezentate mai sus (competențe, activități comunicative, sarcini) ar fi reprezentate pe axa orizontală, urmărind performanța individului în cadrul fiecărei etape în procesul de însușire a limbii. În schimb, axa verticală ar fi rezervată poate celei mai cunoscute clasificări în rândurile specialiștilor în L2, cea privind cele șase niveluri comune de referință, cu ajutorul cărora se poate descrie progresul utilizatorului de la un nivel la altul<sup>34</sup>. Distingem,

<sup>&</sup>lt;sup>26</sup>Angela Bidu-Vrănceanu et alli, *Dicționar de Științe ale limbii (DȘL)*, București, Nemira, 2005, p. 121, *s.v. competență*.

<sup>&</sup>lt;sup>27</sup> CECRL, p. 9.

<sup>&</sup>lt;sup>28</sup> Ibidem.

<sup>&</sup>lt;sup>29</sup> Companion Volume, p. 32.

<sup>&</sup>lt;sup>30</sup> *Ibidem*, p. 33.

<sup>&</sup>lt;sup>31</sup> CECRL, p. 10.

<sup>&</sup>lt;sup>32</sup> *Ibidem*, p. 9.

<sup>&</sup>lt;sup>33</sup> Angela Bidu-Vrănceanu et alli, op. cit. p. 138, s.v. context.

<sup>&</sup>lt;sup>34</sup> După cum e evidențiat și în *Companion Volume*, p. 35.

astfel, un utilizator începător (cu o deosebire între A1, nivelul elementar, si A2, nivel de supravietuire), un utilizator independent (făcând diferentierea între B1, nivelul prag, si B2, nivelul avansat), dar si un utilizator experimentat (C1, nivelul de abilitate operațională efectivă, și C2, nivelul de măiestrie)<sup>35</sup>. Și în această situație e interesant de remarcat, încă o dată, accentul tot mai puternic pus pe utilizator, pe ceea ce acesta poate realiza în actul comunicării, cu mijloacele lingvistice de care dispune la un moment dat, si nu pe ceea ce credem noi că ar trebui să poată face utilizatorul, după cum se evidențiază și în Introductory Guide to the Common European Framework of Reference (CEFR) for English Language Teachers<sup>36</sup>.

Această clasificare a nivelurilor este extrem de importantă în ceea ce privește standardizarea examenelor de competentă lingvistică, centrele de testare<sup>37</sup> adoptând terminologia și propunând teste pe nivelurile corespunzătoare CECRL. Am putea afirma că, de fapt, atât pentru cel care învată limba, cât și pentru cel care o predă/evaluează, aceste niveluri au ajuns să reprezinte o etichetă a conținutului *însusit/evaluat*. Întrucât nivelurile interesează nu numai din perspectiva evaluării, ci si a învătării, căci și în acest act e important să se urmărească parcursul celui care învață o limbă și stadiile de învătare, în CECRL se propun niște niveluri intermediare, între A2, B1 și B2, apărând, astfel, așa-numitele niveluri "plus": A2+, B1+ si B2+. Acestea apar ilustrate si în scalele de descriptori printr-o linie verticală, care delimitează descriptorii necesari primului stadiu dintr-un nivel de aceia care descriu cum se poate parcurge și termina acel nivel<sup>38</sup>. Așadar, vom avea, prin această perspectivă, o separare în trei planuri a însușirii cunoștințelor în cadrul unui nivel.

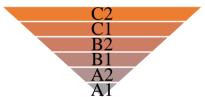


Figura 2 - Nivelurile comune de referință

<sup>&</sup>lt;sup>35</sup> CECRL, p. 23.

<sup>&</sup>lt;sup>36</sup> Introductory Guide to the Common European Framework of Reference (CEFR) for English Language Teachers, Cambridge, Cambridge University Press, 2013, p. 7.

<sup>&</sup>lt;sup>37</sup> De exemplu, în cazul unor centre de examinare cu prestigiu precum Cambridge Assessment English, propuse examenele sunt clasificate, în primul rând. după nivel: http://www.cambridgeenglish.org/exams-and-tests/. De asemenea, DELE (spaniolă), TELC (germană) sau DELF, DALF (franceză).

<sup>&</sup>lt;sup>38</sup> CECRL, pp. 31-32.

Am ales să prezentăm, în continuare, unul dintre cele mai cunoscute instrumente oferite de CECRL. Este vorba despre *Nivelurile comune de referință*. *Scala globală*<sup>39</sup>, unde sunt oferite descrieri de tipul "pot să" pentru toate nivelurile și unde se poate vedea modalitatea în care cunoștințele și abilitățile presupuse pentru fiecare nivel ar trebui să se dezvolte. Pentru a avea o imagine cât mai clară asupra a ceea ce poate face, din punct de vedere lingvistic, un individ la aceste niveluri, am evidențiat prin caractere cursive aspectele definitorii.

Tabelul 1 - Nivelurile comune de referință. Scala globală CECRL

Utilizator	C2	Donto aŭ ŝutologoŭ mugotio făuŭ ofout tot onea en sitegto a
	C2	Poate să înțeleagă, practic fără efort, tot ceea ce citește sau aude.
experimentat		Poate să reconstituie anumite fapte și argumente din diverse surse scrise sau orale, rezumându-le în mod coerent.
		,
		Poate să se exprime spontan, foarte fluent și cu precizie și poate să distingă
	G1	nuanțe fine de sens legate de subiecte complexe.
	C1	Poate să înțeleagă o gamă largă de texte lungi și complicate și să sesizeze semnificațiile implicite.
		Poate <i>să se exprime spontan și fluent</i> , fără a fi obligat aparent să-și caute cuvintele.
		Poate să utilizeze limba cu eficacitate și cu suplețe în viața socială,
		profesională sau academică. Poate să exprime teme complexe, într-un mod clar și bine structurat, demonstrând că stăpânește instrumentele de
		organizare, de structurare și de coeziune a discursului.
Utilizator	B2	Poate să înțeleagă ideile principale din texte complexe pe teme concrete
independent		sau abstracte, inclusiv în discuții tehnice în specialitatea sa.
•		Poate să comunice cu un anumit grad de spontaneitate și de fluență, încât
		conversația cu un locutor nativ să nu presupună efort pentru niciunul
		dintre interlocutori.
		Poate să se exprime clar și detaliat asupra unei game largi de subiecte, poate
		să-și expună <i>opinia asupra unui subiect de actualitate</i> , prezentând avantajele și inconvenientele diferitelor posibilități.
	B1	Poate să înțeleagă <i>elementele esențiale</i> când este folosit <i>un limbaj standard</i>
	DI	clar, pe teme familiare, întâlnite uzual, legate de muncă, școală, timpul
		liber etc.
		Poate să se descurce în majoritatea situațiilor survenite în timpul unei
		călătorii într-o regiune unde este vorbită limba.
		Poate să producă un discurs simplu și coerent pe teme familiare și din domeniile sale de interes.
		Poate să relateze un eveniment, o experiență sau un vis, să descrie o speranță
		sau un scop și să expună, pe scurt, argumentele sau explicațiile unui proiect sau ale unei idei.
Utilizator	A2	Poate să înțeleagă fraze izolate și expresii frecvent folosite în domenii de
începător		interes nemijlocit (de exemplu, <i>informații personale</i> și <i>familiale simple</i> ,
•		cumpărături, mediul înconjurător apropiat, activitatea profesională).
		p. ojestonata).

<sup>&</sup>lt;sup>39</sup> Tabel 1 - Nivelurile comune de referință - Scala globală în CECRL, p. 29, https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000 0168045bc75.

	Poate să comunice în situații simple și obișnuite, care nu solicită decât un schimb de informații simplu și direct, referitoare la subiecte familiare și obișnuite.  Poate să descrie, cu mijloace simple, formația sa profesională, mediul său înconjurător apropiat și să evoce subiecte care corespund nevoilor sale imediate.
A1	Poate să înțeleagă și să utilizeze expresii familiare și cotidiene, precum și enunțuri foarte simple care vizează satisfacerea nevoilor concrete.
	Poate <i>să se prezinte</i> sau <i>să prezinte pe cineva</i> , poate să formuleze și să
	răspundă la întrebări referitoare la detalii personale, ca, de exemplu, unde
	locuiește, persoane cunoscute, lucruri care-i aparțin etc.
	Poate să comunice, într-un mod simplu, cu condiția ca interlocutorul să
	vorbească rar și clar și să fie cooperant.

Din această scală globală se poate observa parcursul unei persoane care învață o limbă străină: de la enunțuri și expresii familiare și cotidiene, pe teme care țin de sfera personală (la nivelul A1), va trece la extinderea comunicării în situații simple, obișnuite (A2), ajungând să se descurce în diferitele situații pe care le poate întâlni în timpul unei călătorii în țara a cărei limbă o învață (B1), până când, spre finalul procesului, va ajunge să înțeleagă tot ce aude și tot ce citește (C2).

Această scală reprezintă, desigur, numai un exemplu, în CECRL fiind propuse "scale pentru toate activitățile de comunicare, cu diferite alte subscale în cadrul fiecărui tip de activitate, unde, cu ajutorul unor descriptori de tip *pot să*, care descriu ce poate face un utilizator în LŢ, se arată încă o dată că accentul cade pe utilizator"<sup>40</sup>. Învățarea centrată pe individ constituie un aspect esențial în CECRL, așa cum foarte convingător o spune și David Little: "nu e niciun accident că, în a doua parte a titlului CECRL: *învățare, predare, evaluare*, învățarea vine înainte de predare, iar predarea, înainte de evaluare. Ordinea celor trei substantive implică o perspectivă centrată pe învățare și pe cel care învață" [tr. n.]<sup>41</sup>. Odată cu apariția *Companion Volume*, descriptorii din aceste scale au fost revizuiți, dar s-au și adăugat scale noi pentru mediere și competență culturală și pluriculturală, precum și pentru limbajul semnelor, instrumente extrem de utile în achizitia oricărei L2.

Este interesant de observat că, în ciuda faptului că CECRL are un rol bine determinat în domeniul însușirii limbilor străine, totuși, cei direct implicați, profesorii și responsabilii de elaborarea materialelor didactice și a curiculelor, de exemplu, nu cunosc decât o parte dintre aspectele conținute de CECRL, în special nivelurile și descriptorii aferenți. North remarca lipsa recunoașterii, în rândul

<sup>&</sup>lt;sup>40</sup> Lavinia-Iunia Vasiu, op. cit., 2020, p. 12.

<sup>&</sup>lt;sup>41</sup> David Little, G. Erickson, "Learner identity, learner agency, and the assessment of language proficiency: Some reflections prompted by the Common European framework of reference for languages", in *Annual Review of Applied Linguistics*, no. 35, 2015, p. 122.

categoriilor mai sus menționate, a abordării adoptate de CECRL sau a schemei descriptive, punând acest fapt și pe seama unei lipse suficiente de preocupare din partea specialiștilor pentru promovarea tuturor beneficiilor aduse de CECRL<sup>42</sup>. Cu toate că unul dintre obiectivele CECRL este acela de a reflecta asupra practicilor din domeniu, atât North, cât și Little și-au exprimat îngrijorarea privind preluarea și aplicarea ca atare a ideilor din CECRL, fără suficientă flexibilitate. Astfel, North susține ideea ca "utilizatorii să împartă sau să combine activitățile și competențele, iar etapele din procesul de învățare să fie adaptate atât contextului local, cât și scopului general" [tr. n.]<sup>43</sup>. Și Little arată că utilizarea parțială a CECRL, fără a reflecta asupra adaptării și a aplicării sale la o limbă specifică face ca documentul să nu mai aibă impactul dorit, mai ales în clasele de L2<sup>44</sup>.

Referindu-se la cazul specialistilor din domeniul limbii române ca L2 privind abordarea CECRL și a terminologiei aduse de acest document, Platon afirmă că "analiza materialelor de predare și evaluare ne relevă cât de puțin am reușit să ne racordăm la tendintele limbilor cu mare traditie în predare"45. Ba mai mult, observă autoarea, "chiar si atunci când se înregistrează o schimbare a ritmului, motivul accelerării pare să țină mai cu seamă de o modă a euro-integrării, decât de constientizarea nevoii de adaptare la principiile elementare ale politicii lingvistice europene"46. De asemenea, se remarcă si rapiditatea cu care a fost acceptată noua repartitie pe niveluri, precum si "căderea în dizgratie a vechii terminologii (începător, intermediar, avansat) în favoarea noilor termeni<sup>47</sup> (A1, A2,....C2). Luând în discutie motivul pentru care au fost atât de usor asimilate și acceptate aceste concepte, autoarea oferă două posibilități: "1. înțelegerea strădaniei specialiștilor de a dispune de aceleasi unităti de măsură și de aceleasi norme; 2. practicienii au sesizat cât de tributari au rămas modalităților intuitive" 48. În același studiu, se face o demonstrație extrem de relevantă în legătură cu pericolul pe care îl poate reprezenta acceptarea ca atare a direcțiilor și a descriptorilor din CECRL, tocmai din cauza

<sup>&</sup>lt;sup>42</sup> Brian North, "The Relevance of the CEFR to Teacher Training", in *Babylonia*, no. 2, 2008, Zürich, p. 55.

<sup>&</sup>lt;sup>43</sup> Brian North, "The CEFR illustrative descriptor scales", in *Modern Language Journal*, no. 91, 2007, pp. 656–659.

<sup>&</sup>lt;sup>44</sup> David Little, "The Common European Framework of Reference for Languages: A research agenda.", in *Language Teaching*, no. 44, 2011, p. 383 *apud* Ioana Sonea, *op. cit.*, 2018, p. 85.

<sup>&</sup>lt;sup>45</sup> Elena Platon, "RLS - Validarea competențelor lingvistice în conformitate cu Cadrul European Comun de Referință și cu Pașaportul lingvistic European", în vol. *Romània Orientale*, no. XXI, Roma, Bagatto Libri, 2008, p. 306.

<sup>&</sup>lt;sup>46</sup> Elena Platon, op. cit., 2008, p. 306-307.

<sup>&</sup>lt;sup>47</sup> *Ibidem*, p. 307.

<sup>&</sup>lt;sup>48</sup> *Ibidem*, pp. 307-308.

caracterului lor general, în care pot fi incluse mostre de limbă de nivel inferior, cât și superior<sup>49</sup>. De exemplu, o formulare de tipul *propoziție simplă* este una foarte generală, sub care, din perspectiva vorbitorului nativ și neavând alte instrumente de măsurare la dispoziție, am fi tentați să includem enunțuri pe care, din perspectiva L2, le plasăm pe niveluri diferite, de exemplu: "1. *Ea este Elena*. (A1); 2. *O cheamă Elena*. (A2); 3. *Nu știam că o cheamă Elena*. (B1); 4. *Numele celei mai bune prietene a mele e Elena*. (B2)"<sup>50</sup>. Aceeași situație confuză este identificată și în cazul modului de formulare a cerințelor sau al selecției textelor de utilizat ca input în cadrul activităților de comunicare.

Un aspect care i se impută CECRL frecvent se referă la lipsa validării empirice a descriptorilor si a corelării acestora cu cercetarea din domeniul achizitiei limbii a doua, situație cunoscută (și recunoscută) chiar de către autorii CECRL. Bunăoară, însuși North recunoaște acest minus al materialului, însă justifică și sustine metoda folosită, deoarece, în momentul elaborării CECRL, astfel de cercetări si descrieri nu existau<sup>51</sup>. Si Alderson aduce aceleasi critici, evidentiind nevoia de cercetări mai elaborate în ceea ce priveste competenta lingvistică în principalele limbi europene. Mai mult, acesta consideră că ar fi necesar un corpus al nonnativilor (learner corpus), bazat pe o evaluare individuală a acestora și abia apoi pe rezultate comune<sup>52</sup>. La un deceniu după observațiile lui North și Alderson, Wisniewski confirmă și reafirmă necesitatea unei analize empirice, bazată pe un corpus al nonnativilor, pentru a valida descriptorii din CECRL. Astfel, autoarea sustine că scalele ar trebui legate de modele ale abilității de comunicare lingvistică, să se afle în acord cu cercetarea din domeniul achiziției limbii a doua, dar și să poată fi relationate la mostre de L2 (empirical learner language). Mai mult, întăreste cele spuse printr-o afirmație, considerăm noi, extrem de relevantă: "e important să se ofere dovezi că evaluatorii umani pot implementa aceste scale" [tr. n.]<sup>53</sup>. Astfel, încă o dată se remarcă necesitatea validării și a descrierii nivelurilor de referință din CECRL prin raportare la situația particulară a fiecărei limbi.

\_

<sup>&</sup>lt;sup>49</sup> Observații similare pot fi regăsite și în Elena Platon, "Descrierea minimală a limbii române (A1, A2, B1, B2) – un instrument lingvistic necesar în domeniul RLS/RLM, în Elena Platon, Ioana Sonea, Lavinia Vasiu, Dina Vîlcu, *Descrierea minimală a limbii române. A1, A2, B1, B2*, Cluj-Napoca, Editura Casa Cărții de Știință, 2014, pp. 3-4.

<sup>&</sup>lt;sup>50</sup> Elena Platon, op. cit., 2008, p. 309.

<sup>&</sup>lt;sup>51</sup> Sonea, Ioana, *op. cit.*, 2018, p. 88. și Brian North, *op. cit.*, 2007, p. 657.

<sup>&</sup>lt;sup>52</sup> C. J. Alderson, "The CEFR and the need for more research" in *Modern Language Journal*, no. 91, 2007, p. 661.

<sup>&</sup>lt;sup>53</sup> K. Wisniewski, "Empirical learner language and the levels of the Common European Framework of Reference" in *Language Learning*, 67(S1), 2017, p. 9.

Cu toate că North sustine că cele sase niveluri comune de referintă reprezintă percepția profesorului asupra abilităților lingvistice ale studentului și nu constituie niste etape în procesul de dezvoltare, de-a lungul timpului, a apărut întrebarea dacă nu cumva cele sase niveluri sunt, de fapt, niște etape logice în procesul de învățare a unei limbi<sup>54</sup>. Asadar, există o serie de specialisti care se întreabă dacă un individ care învață o limbă străină trebuie obligatoriu să parcurgă, în ordine, în procesul său de achizitie a limbii, nivelurile A1-A2-B1-B2-C1-C2 sau dacă un individ, care se află la nivelul B1, de exemplu, a trecut obligatoriu prin nivelurile A1-A2 sau dacă e posibil să fi aiuns la acest nivel (B1) fără a fi parcurs A1 si A2<sup>55</sup>. Un răspuns la aceste întrebări poate veni din Companion Volume unde se subliniază că scalele din CECRL au fost elaborate pentru "a realiza un profil al abilității unui individ" [tr. n.]<sup>56</sup>. Mai mult, se arată că e prea puțin credibil ca toți utilizatorii (care au fost plasați global, la nivelul B1, de exemplu) să fie capabili să facă exact aceleași lucruri definite la B1, în toți descriptorii și în toate scalele CECRL. "E mai probabil că nivelul lor general e de fapt A2 sau A2+ în relatie cu unele activităti comunicative, si B1 sau B2, în relatie cu altele, în functie de competentele lor generale, de vârstă, experientă etc." [tr. n.]<sup>57</sup>.

În contextul acestei discuții, Lowie propune perspectiva teoriei sistemelor dinamice în învățarea limbilor. Conform acestei teorii, "un sistem dinamic e repetitiv, cu diverse stadii de dezvoltare, fiecare dintre acestea fiind influențate atât prin factori externi, cât și interni. Prin urmare, sistemele dinamice sunt imprevizibile, iar fiecare traiectorie de dezvoltare este cu precădere individuală [tr. n.]"58. Sonea atrage atenția asupra implicațiilor pe care un astfel de sistem le-ar aduce procesului educațional<sup>59</sup>. Astfel, în situația descrisă de către Lowie, planificarea cursurilor și curriculumul nu și-ar mai găsi locul, ceea ce ar conduce spre un proces de învățare "imprevizibil, variabil și individual" [tr. n.]<sup>60</sup>, ceea ce ar fi extrem de dificil, dacă nu chiar imposibil, de a fi integrat în sistemul de educație formală. Din acest motiv se impune să abordăm situația profilurilor *egale (flat profiles)* și a acelora *inegale (uneven profiles)*, atunci când se discută despre nivelurile comune de referință din

<sup>&</sup>lt;sup>54</sup> Ioana Sonea, *op. cit.*, 2018, p. 91 şi W. M. Lowie, *The CEFR and the dynamics of second language learning: Trends and challenges. Language Learning in Higher Education. Journal of the European Confederation of Language Centres in Higher Education* (CercleS), 2(1), 2012, Berlin/Boston, De Gruyter, p.18, disponibilă online aici: https://doi.org/10.1515/cercles-2012-0002.

<sup>&</sup>lt;sup>55</sup> *Ibidem* și J. H. Hulstijn, "The common European framework of reference for languages. A challenge for applied linguistics" in *International Journal of Applied Linguistics*, no. 165, 2013, p. 10.

<sup>&</sup>lt;sup>56</sup> Companion Volume, p. 53.

<sup>&</sup>lt;sup>57</sup> *Ibidem*, p. 53.

<sup>&</sup>lt;sup>58</sup> W. M. Lowie, op. cit., 2012, p. 19.

<sup>&</sup>lt;sup>59</sup> Ioana Sonea, *op. cit.*, 2018, p. 91.

<sup>60</sup> W. M. Lowie, op. cit., 2012, p. 22.

CECRL<sup>61</sup>. Autori precum Huljistin sau North iau în discuție această temă, susținând că "profilurile inegale sunt regula, iar cele egale sunt excepția" [tr. n.]<sup>62</sup> sau că "o etichetă precum «A2» reprezintă doar o metodă convenientă de a denumi un profil complex" [tr. n.]<sup>63</sup>.

În ceea ce privește limba română, putem remarca un interes crescând pentru CECRL, în special în ceea ce privește nivelurile comune de referință și activitățile comunicative. Dintre materialele dedicate temei, se evidențiază seria de articole aparținând lui Platon, axate pe necesitatea de a adapta direcțiile și conceptele din CECRL la specificul limbii române<sup>64</sup>, iar, odată cu teza de doctorat a Laviniei-Iunia Vasiu, intitulată *Achiziția limbii române ca L2. Interlimba la nivelul A1*, s-a făcut o prezentare exhaustivă a principalelor aspecte din jurul problematicii CECRL, motiv pentru care nu mai detaliem și alte aspecte legate de acest document.

După cum am arătat, valoarea CECRL în domeniul politicilor lingvistice este de necontestat. Totuși, câteva aspecte din CECRL necesită cercetare suplimentară. Pe de-o parte, vorbim despre o "cercetare empirică în ceea ce privește nivelurile comune de referință din CECRL, în special bazată pe un corpus al vorbitorilor nonnativi (*learner corpus*), iar, pe de altă parte, e necesar să se observe felul în care se relaționează nivelurile la domeniul de achiziție a L2" [tr. n.]<sup>65</sup>.

# 1.2. Cadrul european comun de referință pentru limbi: învățare, predare, evaluare. Companion Volume

Având în vedere perioada îndelungată scursă de la apariția CECRL, precum și criticile firești aduse acestuia<sup>66</sup>, în anul 2018 a apărut, tot sub îngrijirea Consiliului Europei, *Companion Volume with New Descriptors*<sup>67</sup>, urmat în 2020, de versiunea

<sup>63</sup> Brian North, Putting the Common European Framework of Reference to good use. Language Teaching, 47(02), 2014, p. 230.

<sup>66</sup> Pentru o prezentare a celor mai importante critici aduse CECRL, vezi Lavinia-Iunia Vasiu, *op.cit*, 2020, pp. 27-33.

<sup>&</sup>lt;sup>61</sup> Pentru o discuție în profunzime pe tema profilurilor egale sau inegale, vezi Ioana Sonea, *op. cit.*, 2018, pp. 91-93.

<sup>&</sup>lt;sup>62</sup> J. H. Hulstijn, op. cit., 2013, pp. 10–11.

<sup>&</sup>lt;sup>64</sup> Semnalăm, de asemenea, și funcționarea masteratului *Româna ca limbă străină/nematernă*, precum și a unui modul opțional de RLS la nivelul masterat, la Facultatea de Litere de la Universitatea Babeș-Bolyai, care are în componență cursuri consacrate aceleiași problematici, susținute de către conf. univ. dr. Elena Platon.

<sup>65</sup> Ioana Sonea, op. cit., 2018, p.106.

<sup>&</sup>lt;sup>67</sup> Companion Volume, versiunea 2018, disponibil aici: https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989.

finală, numită Companion Volume<sup>68</sup>. Conform paginii oficiale a proiectului<sup>69</sup>, volumul trebuie văzut ca un material complementar la CECRL. Acest document este extrem de important întrucât aduce descrieri pentru elemente esentiale care lipseau din materialul din anul 2001. Ne referim aici la descriptorii pentru mediere și competentă culturală și pluriculturală, precum și la descrieri clare pentru nivelurile "plus" din CECRL sau chiar și pentru un presupus nivel preA1. Pe lângă aceste aspecte, Companion Volume aduce explicatii mai elaborate pentru nivelurile A1, C1 si C2. După cum însisi autorii afirmă în cuvântul-înainte al lucrării<sup>70</sup>, acest material reprezintă o continuare a CECRL, ca urmare a cererilor venite din rândul utilizatorilor de a dezvolta descriptorii ilustrativi și de a-i completa cu unii pe mediere si interactiune online, de a produce versiuni ale acestora pentru tinerii utilizatori (young learners) și pentru limbajul semnelor, precum și de a se concentra pe o descriere mai completă a nivelurilor A1 și C. Astfel, cele mai semnificative schimbări au avut loc în cazul medierii, al competentei plurilingvistice/pluriculturale, dar si al limbajului semnelor, unde nu au existat alti descriptori<sup>71</sup>. Decizia autorilor de a acoperi si aceste domenii ni se pare salutară si considerăm întreg demersul parcurs extrem de util, chiar dacă nu (exact) acesta reprezintă obiectul cercetării noastre<sup>72</sup>.

Nu sunt de neglijat însă nici celelalte adăugiri aduse, în ceea ce privește nivelurile. După cum am subliniat anterior, încă din CECRL, s-a discutat și despre un set de niveluri intermediare<sup>73</sup>. În *Companion Volume* se reia discuția, iar acestea apar mai bine dezvoltate și, în plus, apare, din nou, un nivel "preA1"<sup>74</sup>. Înainte de a atinge cel mai de jos nivel de performanță lingvistică într-o limbă, odată cu *Companion Volume*, s-a reluat ideea din CECRL că "ar putea exista o gamă de sarcini specifice, relevante pentru nevoile lor, pe care cei care învață o limbă le pot

<sup>&</sup>lt;sup>68</sup> Companion Volume, versiunea 2020, <a href="https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4">https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4</a>. Întrucât la momentul redactării lucrării noastre ne-am bazat pe volumul publicat în 2018, precizăm că referințele bibliografice la Companion Volume au în vedere volumul publicat în 2018.

<sup>&</sup>lt;sup>69</sup>https://www.coe.int/en/web/education/-/the-cefr-companion-volume-with-new-descriptors-is-now-available-online-?desktop=true.

<sup>&</sup>lt;sup>70</sup> Companion Volume, pp. 21-22.

<sup>&</sup>lt;sup>71</sup> Ibidem.

<sup>&</sup>lt;sup>72</sup> Singura aplicabilitate a competenței pluriculturale în cazul nostru poate fi în cazul necesității de a explica cazurile de însuşire mai rapidă sau mai lentă a unor elemente lexicale, proces privit prin prisma cunoașterii deja a mai multor limbi.

<sup>&</sup>lt;sup>73</sup> CECRL, pp. 32-36.

<sup>&</sup>lt;sup>74</sup> E interesant de observat că există, de exemplu, la Cambridge Assessment English teste pentru copii, la nivelul preA1 – *Pre A1 Starters*.

face în mod eficient folosind o gamă restrânsă de structuri" [tr. n.]<sup>75</sup>. Acest nivel e luat în considerare mai ales în cazul copiilor și reprezintă rezultatul a două proiecte (Swiss Lingualevel, Japanese CEFR\_J) care au avut ca grup-țintă elevi din ciclurile primar și secundar. Acesta e considerat ca o "piatră de hotar în drumul cuiva spre nivelul A1, fiind momentul în care individul are un repertoriu de cuvinte și formule **memorate,** pe care le folosește când e imperios necesar, dar când încă nu are capacitatea de a **genera** discurs" [tr. n.]<sup>76</sup>. Considerăm extrem de utilă reluarea discuției și introducerea acestor noi descriptori, întrucât, în acest fel, se vor putea urmări și micile progrese făcute de către cursanți, chiar și în ceea ce privește atingerea acestui nivel preA1, care, chiar și în rândul adulților<sup>77</sup>, ar putea constitui un factor motivant pentru învătare.

În permanență, pe parcursul volumului, se accentuează ideea că adăugirile aduse prin *Companion Volume* nu afectează constructul descris în CECRL, ci reprezintă o dezvoltare, o continuare a ideilor enunțate acolo. Din punctul nostru de vedere, *Companion Volume* este un material mult mai simplu de parcurs decât CECRL, grație utilizării unei forme de prezentare a informației mult mai prietenoasă cu cititorul – casete, diagrame, chiar explicații de ordin istoric ale unor concepte etc.

Odată cu *Companion Volume*, se accentuează preferința pentru colaborare în comunicare: propunându-se un alt model, fundamentat pe interacțiune, unde are loc construirea sensului. Astfel, "activitățile comunicative sunt prezentate sub forma a patru tipare de comunicare: receptare, producere, interacțiune și mediere" [tr. n.]<sup>78</sup>. Conform *Companion Volume*<sup>79</sup>, "activitățile de mediere, atât în receptare, cât și în producere (scrisă sau orală), fac posibilă comunicarea între două persoane care din diferite motive nu sunt capabile să comunice direct. Medierea se bazează pe activități de traducere, de interpretare sau de parafrazare, prin care se asigură o reformulare a textului-sursă astfel încât să fie accesibil unui alt actant în procesul comunicării" [tr. n.]<sup>80</sup>.

Adoptarea acestor tipare a fost argumentată de către North într-un studiu publicat în 1997, unde se evidenția faptul că propunerea acestor categorii (receptare, producere, interacțiune și mediere) este justificată prin faptul că ilustrează mult mai realist felul în care oamenii utilizează cunoștințele lingvistice în comunicare,

<sup>&</sup>lt;sup>75</sup> Companion Volume, p. 35.

<sup>&</sup>lt;sup>76</sup> *Ibidem*, p. 46.

<sup>&</sup>lt;sup>77</sup> Ne referim, aici, de exemplu, la participanții la cursurile de vară sau la alte tipuri de cursuri de limba română ori la atât de actuala problematică a integrării lingvistice a migranților (vezi pentru mai multe detalii site-ul Consiliului Europei: https://www.coe.int/en/web/language-policy/home#P30\_2633).

<sup>&</sup>lt;sup>78</sup> Companion Volume, p. 30.

<sup>&</sup>lt;sup>79</sup> Vezi Figura 2, din *Companion Volume*, p. 32.

<sup>80</sup> Companion Volume, p. 32.

susținând, în continuare, și conexiunea mult mai facilă între limbajul general și cel specializat sau, pur și simplu, în ceea ce privește activitățile pedagogice, care pot avea un caracter mult mai autentic, asemănându-se unei "felii din viață"<sup>81</sup>. Totuși, e de remarcat faptul că, până la *Companion Volume*, unde medierea ocupă un loc extrem de important, deși a fost inclusă în CECRL, aceasta a fost mai degrabă neglijată<sup>82</sup>.

După cum am evidențiat deja în câteva rânduri, CECRL a adus o multitudine de schimbări în domeniul limbilor străine, având un impact major atât în evaluare (standardizarea testelor de limbi), cât și în predare (elaborarea de metodologii, curricule, syllabusuri și diferite materiale necesare în predare), marcând momentul în care "competența lingvistică nu mai e definită doar în termeni gramaticali și lexicali, ci pe niveluri de stăpânire diferite calitativ și cantitativ"83.

## 1.3. Materiale elaborate pornind de la CECRL

În procesul firesc de adoptare și aplicare al oricărui material de o asemenea anvergură, au urmat o mulțime de materiale care au încercat să pună în aplicare CECRL, exemple în acest sens fiind *Portofoliul lingvistic european* sau proiectul *DIALANG*, asupra cărora nu insistăm aici<sup>84</sup>. Un alt exemplu este volumul intitulat *Relating Language Examinations to the Common European Framework of Reference for Languages: Teaching, Learning and Assessing (CEFR). A Manual*<sup>85</sup>, care are rolul de a-i ghida pe furnizorii de teste în procesele de dezvoltare, de administrare și de continuă îmbunătățire a testelor în scopul relaționării lor la CECRL. Acestui material i se alătură alte două volume, gândite a fi folosite împreună cu CECRL și cu acest *Manual*, e vorba despre *Manual for Language Test Development and Examining*<sup>86</sup> și *Guidelines for Development of Language for Specific Purposes Tests*<sup>87</sup>, ambele produse de ALTE (Association of Language

<sup>&</sup>lt;sup>81</sup> După Brian North, *Perspectives on Language Proficiency and Aspects of Competence: a reference paper defining categories and levels*, Strasbourg, Council of Europe CC-Lang (94)20, 1997.

<sup>82</sup> În motivația autorilor privind nedezvoltarea în CECRL a acestei teme stă dimensiunea deja extinsă a materialului (250 de pagini).

<sup>83</sup> Lavinia-Iunia Vasiu, op. cit., 2020, pp. 27-28.

<sup>&</sup>lt;sup>84</sup> Pentru mai multe detalii, vezi Lavinia-Iunia Vasiu, op. cit., 2020, p. 30.

<sup>&</sup>lt;sup>85</sup> Council of Europe, Relating Language Examinations to the Common European Framework of Reference for Languages: Teaching, Learning and Assessing (CEFR). A Manual, Language Policy Division, Strasbourg, 2009.

<sup>&</sup>lt;sup>86</sup> Disponibil online aici: https://rm.coe.int/manual-for-language-test-development-and-examining-for-use-with-the-ce/1680667a2b.

<sup>&</sup>lt;sup>87</sup> Disponibil online aici:

https://www.alte.org/resources/Documents/6093%20LSP%20Supplement%20-%20WEB.pdf.

Testers in Europe). Acestea se diferențiază de primul manual prin faptul că sunt mult mai detaliate și ating subiecte care în materialul din 2009 nu fuseseră (suficient) abordate (vezi Manual for Language...) sau care descriu trăsături caracteristice testelor pentru limbajul specializat, cum este cazul materialului din 2018.

Tot de dată mai recentă este proiectul Reference Level Descriptions – RLD<sup>s</sup> for national and regional languages<sup>88</sup>, care, urmând îndemnul ca utilizatorii "să adapteze utilizarea CECRL pentru a se potrivi limbii cu care acestia lucrează, precum și contextului lor" [tr. n.]89, propune un plan-cadru pentru descrieri ale limbilor pe nivelurile comune de referință. Acestea sunt, de fapt, descrieri-cadru, pe niveluri, realizate pentru diferite limbi (de către echipe de lingvisti din acele tări, cu metodele proprii de analiză, dispunând de ajutorul expertilor din Consiliul Europei), prin care se adaptează descriptorii pe niveluri limbii în cauză. În general, acestea poartă numele de "profil al limbii" [tr. n.]<sup>90</sup>. Astfel, sunt descrise și elaborate elementele gramaticale si lexicale necesare pentru dobândirea tuturor competentelor, pe fiecare nivel.

Trebuie remarcat aici că tocmai acestea constituie, în esentă, elementele care i se reprosaseră CECRL: modul de descriere a nivelurilor si necesitatea de validare empirică a descrierilor pentru nivelurile comune de referință.

Conform paginii oficiale a Consiliului Europei<sup>91</sup>, unde există mai multe informatii despre aceste descrieri, există deja descrieri pentru 11 limbi: germană, engleză, spaniolă, franceză, italiană, croată, cehă, georgiană, lituaniană, portugheză și turcă. În momentul în care se decide dezvoltarea unei descrieri de referintă (RLD) într-o limbă, se poate opta pentru o descriere completă pe un nivel, pe o serie de niveluri, precum și pe o "feliuță", atâta vreme cât e clar delimitat scopul acestei descrieri" [tr. n.]<sup>92</sup>. Indiferent de opțiune, este recomandat ca autorii să respecte indicațiile privind metodologia de elaborare a descrierilor pe niveluri de competență lingvistică a limbilor (aici, incluzându-se aspectele de ordin fonetic, lexical, gramatical, si nu numai), precum si utilizarea unui corpus cât mai divers. În scopul asigurării unei perspective științifice a documentului rezultat, în ghidul pentru elaborarea RLD, se indică folosirea următoarelor materiale în alcătuirea corpusului: liste de frecvență, un set de texte de mari dimensiuni (nu neapărat literare), fragmente

<sup>88</sup> Informatii aici: https://www.coe.int/en/web/common-european-framework-referencelanguages/reference-level-descriptions#26120048\_26120993\_True.

<sup>&</sup>lt;sup>89</sup> Using the CEFR Principles of Good Practice, October 2011, Cambridge ESOL, Cambridge, p. 4.

<sup>90</sup> De exemplu, English Profile, Profile Deutsch, Profilio della lingua italiana etc.

<sup>91</sup> www.coe.int/t/dg4/linguistic/dnr/ EN.asp.

<sup>&</sup>lt;sup>92</sup> Using the CEFR Principles of Good Practice, October 2011, Cambridge ESOL, Cambridge, p. 24.

din materialele care asigură achiziția limbii, atât în cadru instituționalizat, cât și întrunul natural, producții (texte) ale candidaților, în special, producții elaborate în cadrul evaluărilor, răspunsuri intuitive și rapide din partea unor informatori instruiti<sup>93</sup>.

Nu în ultimul rând, se cuvine să amintim și materialele destinate exclusiv predării, elaborate de EAQUALS, *A Core Inventory for General English*<sup>94</sup> și *Inventaire linguistique des contenus clés aux niveaux du CECRL*<sup>95</sup>, cu rol deosebit în predare, așa cum se declară chiar pe site-ul EAQUALS: "conțin aspecte care pot ghida profesorul în scopul unei predări adecvate la fiecare nivel din CECRL"<sup>96</sup>. O prezentare mai detaliată a *Core Inventory* am realizat în volumul coordonat de Platon, *Evaluarea limbii române la ciclul primar: EVRO P5*<sup>97</sup>.

<sup>&</sup>lt;sup>93</sup> Criteriile sunt preluate din *Reference Level Descriptions for National and Regional Languages* (*RLD*). *Guide for the production of RLD*, Version 2, Language Policy Division, Council of Europe, Strasbourg, November 2005, p. 6.

<sup>&</sup>lt;sup>94</sup> Brian North, Angeles Ortega, Susan Sheehan, A Core Inventory for General English, Eaquals, 2014.

<sup>&</sup>lt;sup>95</sup> Brian North, *Inventaire linguistiquev des contenus clés aux niveaux du CECRL*, Eaquals, 2015. <sup>96</sup> https://www.eaquals.org/our-expertise/cefr/our-work-practical-resources-for-language-teaching/.

<sup>&</sup>lt;sup>97</sup> Elena Platon (coord.), Antonela Arieșan, Luminița Fodor, *Evaluarea limbii române la ciclul primar: EVRO P5*, Cluj-Napoca, Editura Casa Cărții de Știință, 2015.

### 1.4. Competența lexicală

Așa cum se arată în *Companion Volume*, încă de la începutul anilor 1980, modelele de comunicare din lingvistica aplicată au influențat CECRL. Indiferent la care model facem referire, toate au în comun patru aspecte esențiale: competența strategică, competența lingvistică, competența sociolingvistică/socioculturală și cea pragmatică<sup>98</sup>. Competența strategică e tratată în relație cu activitățile de comunicare, ceea ce face ca *noua schemă a competențelor lingvistice de comunicare* (*Figura 3*) să fie axată pe celelalte 3 direcții, care în timpul utilizării limbii sunt întotdeauna întrepătrunse<sup>99</sup>.

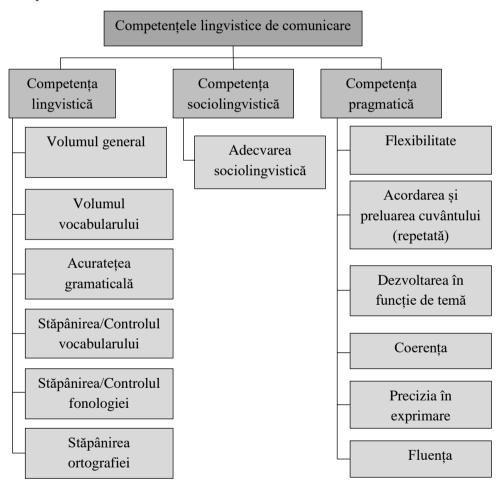


Figura 3 - Noua schemă descriptivă a competențelor lingvistice de comunicare

39

<sup>98</sup> Companion Volume, p. 130.

<sup>&</sup>lt;sup>99</sup> Ibidem.

Conform schemei descriptive, competența lexicală (*lexical competence*) este parte a competenței lingvistice, fiind definită, în CECRL, drept "cunoașterea și capacitatea de a folosi vocabularul unei limbi"<sup>100</sup>. Aceasta e formată din elemente lexicale și gramaticale, prezentate în continuare după accepțiunea CECRL. Astfel, prin elemente lexicale, conform cadrului, se înțeleg următoarele tipuri de elemente<sup>101</sup>:

• *"expresii fixe*, formate din mai multe cuvinte, predate și învățate ca atare"<sup>102</sup>. Aici sunt incluse, după CECRL: 1. *formulele memorate*, care pot sau nu face parte din categoria funcțiilor comunicative. De exemplu, formule de salut, urări, proverbe etc.; 2. *expresii, metafore* de tipul *albă ca zăpada*; 3. *structuri fixe*, în care se pot insera alte elemente pentru a forma conținut. De pildă: *Vă rog..., Aș dori..., Cât costă..., Ce doriți*? etc., 4. *locuțiuni sau expresii frazeologice* care, de obicei, sunt cunoscute ca apărând împreună. De exemplu, *în fața casei, în centrul orașului, a face duș, a da telefon* etc.

#### • cuvinte izolate.

Elementele gramaticale, pe de altă parte, sunt cele deja stabilite spre a fi studiate la fiecare nivel prin descrieri/curricule, în cazul fiecărei limbi. De exemplu, în ceea ce privește limba română, aceste informații se găsesc în *Descrierea minimală a limbii române. A1, A2, B1, B2*<sup>103</sup>. Desigur că atât elementele lexicale, cât și cele gramaticale, propuse aici și repartizate pe niveluri, au numai un caracter orientativ<sup>104</sup>, până la momentul în care acestea vor fi confirmate sau adaptate prin analize empirice, bazate pe corpusuri ale nonnativilor.

După cum e deja cunoscut, principala dovadă a achiziției unei limbi străine stă în progres, adică în faptul că în discursul vorbitorului reușesc să pătrundă noi forme, pe care acesta încearcă să le folosească. Deși am crede că progresul ar presupune dovedirea unui grad de măiestrie în utilizarea acestor noi structuri, în *Companion Volume* se arată că important e tocmai faptul că individul "încearcă să folosească structuri lingvistice mai complexe, asumându-și riscuri și ieșind din zona sa de confort din punct de vedere lingvistic. Toate aceste etape reprezintă, de fapt, o parte esențială din procesul de învățare" [tr. n.]<sup>105</sup>. Așadar, studenții vor avea tendința de a arăta un control limitat asupra unor structuri morfosintactice și lexico-semantice

<sup>&</sup>lt;sup>100</sup> CECRL, p. 110.

<sup>&</sup>lt;sup>101</sup> *Ibidem*, pp. 110-111.

<sup>&</sup>lt;sup>102</sup> *Ibidem*, p. 110.

<sup>&</sup>lt;sup>103</sup> Elena Platon, Ioana Sonea, Lavinia Vasiu, Dina Vîlcu, op. cit., 2014.

<sup>&</sup>lt;sup>104</sup> Elena Platon, op. cit., 2014, p. 9.

<sup>&</sup>lt;sup>105</sup> Companion Volume, p. 131.

nou învățate tocmai datorită dorinței lor de a se autodepăși, de a progresa și de a-și însuși limba. Iar acest comportament e important de identificat și de evaluat obiectiv. În astfel de situații, e important de văzut mesajul, în ansamblul său, și nu doar din perspectiva acurateței, cum de cele mai multe ori avem tendința de a o face, probabil din cauza faptului că reprezintă palierul cel mai simplu și mai comod de controlat și de evaluat în acest moment.

Conform CECRL, descriptorii care urmăresc progresul unui utilizator pot fi folosiți, pe de-o parte, în procesul individual de (auto)evaluare, iar, pe de altă parte, în general, pentru a ilustra progresul și pentru a crea niște repere pentru fiecare nivel. În evaluare, după CECRL<sup>106</sup>, acești descriptori pot să apară sub trei forme:

- a. scale (*scale*) "de obicei, combinând descriptorii pentru diferite categorii în cadrul unui singur paragraf, cu caracter general, pentru fiecare nivel. Această prezentare a descriptorilor e și cea mai comună, de altfel" [tr. n.]<sup>107</sup>. După cum am arătat deja, CECRL propune propune o multitudine de scale.
- b. liste de control (*checklist*) de obicei, "o listă de control fiind propusă pentru fiecare nivel relevant, cu descriptori grupați sub diferite categorii" [tr. n.]<sup>108</sup>.
   Aceste liste de control sunt destul de frecvente, mai ales în cadrul autoevaluării, și sunt destul de rar folosite în evaluările față-n față.
- c. grile, axate pe anumite categorii (*grid*) de fapt, "un set de scale paralele pentru categorii separate (criterii)" [tr. n.]<sup>109</sup>. Cu ajutorul acestui tip de instrument se pot realiza profiluri de diagnostic (stabilirea profilului unui utilizator, de exemplu). Totuși, dezavantajul acestei grile, potrivit CECRL, este că selectează un număr limitat de criterii<sup>110</sup>.

De reținut este că, indiferent de forma în care sunt regăsibili acești descriptori, ei reprezintă numai niște repere generale. Și Platon arată că scalele din CECRL oferă "doar niște linii directoare, impunându-se ca listele detaliate ale microfuncțiilor, ale formelor gramaticale și ale vocabularului să fie prezentate în formulările lingvistice specifice fiecărei limbi" [tr. n.]<sup>111</sup>.

Prezentăm, în continuare, cele trei scale relevante pentru cercetarea noastră: scala lingvistică generală, cea pentru volumul vocabularului și cea pentru controlul vocabularului. Scalele au fost preluate și traduse din CECRL și *Companion Volume*,

<sup>&</sup>lt;sup>106</sup> CECRL, p. 181.

<sup>&</sup>lt;sup>107</sup> Ibidem.

<sup>&</sup>lt;sup>108</sup> Ibidem.

<sup>109</sup> Ibidem.

<sup>&</sup>lt;sup>110</sup> Ibidem.

<sup>&</sup>lt;sup>111</sup> Elena Platon, op. cit., 2008, p. 309.

versiunile în limba engleză<sup>112</sup>. Cu toate că am marcat principalele trăsături pentru toate nivelurile din cele 3 scale luate în discuție, în continuare, în analiza noastră ne vom îndrepta atenția asupra descriptorilor disponibili pentru nivelurile preA1, A1 și A2.

O observație care se impune a fi făcută cu referire la toate cele trei scale e cea privitoare la lipsa unor descriptori la anumite niveluri inferioare. În aceste cazuri, prin *nu există* înțelegem că nu a apărut un descriptor în grila din CECRL, dar e disponibil în cea din *Companion Volume*, în vreme ce prin *nu există descriptor* înțelegem că nu este inclus niciun descriptor în acest moment în niciunul dintre cele două documente, ceea ce, conform autorilor *Companion Volume*, nu înseamnă neapărat că nu există abilitatea în discuție la acel nivel, ci doar că nu s-au găsit suficiente dovezi/exemple la momentul redactării materialului pentru a propune un descriptor în acele cazuri. Desigur că, și în această situație, doar o analiză empirică ne poate arăta dacă un descriptor pentru stăpânirea vocabularului la A1, bunăoară, este de luat în calcul sau, dacă, din contră, nu putem vorbi, într-adevăr, despre un control al vocabularului la acest nivel.

Dintr-o succintă analiză a descriptorilor din CECRL, respectiv *Companion Volume*, putem ușor observa că, în general, descrierile sunt păstrate în cel mai recent material, existând totuși câțiva noi descriptori pe care i-am marcat diferit în scale. O noutate este aceea că, odată cu noii descriptori, la unele niveluri, se oferă mai multe detalii, pe de-o parte, iar, pe de altă parte, apare și distincția între ceea ce poate recepta și ceea ce poate produce un vorbitor. De exemplu, la nivelul A1, în *Scala lingvistică generală*<sup>113</sup> se face distincția între "deține un repertoriu de bază" (cunoștințe mai degrabă receptive) și "poate utiliza" (cunoștințe productive). De reținut e descriptorul aferent nivelului preA1 ("Poate folosi *cuvinte izolate* și *expresii de bază* pentru a oferi *informații simple despre sine*."), separându-se, astfel, vechiul descriptor de A1 și făcându-se diferența între *cunoștințe despre sine* și *despre situațiile cotidiene*, care, până la *Companion Volume*, erau abordate împreună.

Scala lingvistică generală operaționalizează o serie de concepte. Vorbim, astfel, despre: 1. **cadru** (*setting*), care devine tot mai larg. Se pleacă de la *sine* (preA1) și *situații concrete imediate* (A1) pentru a se dezvolta în *situații cotidiene*, *predictibile*, la A2, și pentru a se ajunge la *situații imprevizibile*, *pe teme comune*, la B1, până la *orice cadru*, la C2; 2. *tipul structurilor lexico-gramaticale*, care se dezvoltă în același fel în scală, de la *cuvinte izolate* (preA1), la *fraze memorate* (A1)

<sup>&</sup>lt;sup>112</sup> Companion Volume, pp. 131, 132, 134.

<sup>&</sup>lt;sup>113</sup> *Ibidem*, p. 131.

până la "structuri complexe și variate pentru a exprima gânduri complexe, a sublinia sau a elimina ambiguitatea" [tr. n.]<sup>114</sup>; 3. **limitări**, reprezentate prin pauze sau ezitări, la început, evidente, reperabile, până când se ajunge în situația ca utilizatorul să nu arate în niciun fel că e nevoit să reducă ceea ce vrea să spună (la C2)<sup>115</sup>.

Tabelul 2 - Scala lingvistică generală din CECRL

Nivel	CECRL	Companion Volume
C2	Poate exploata cu măiestrie și cu	Poate exploata cu măiestrie și cu stăpânire
	stăpânire deplină o gamă foarte largă	deplină o gamă foarte largă de discursuri pentru
	de discursuri pentru a-și formula cu	a-și formula <i>cu precizie opiniile</i> , pentru <i>a</i>
	precizie opiniile, pentru a accentua, a discrimina și a înlătura ambiguitatea.	accentua, petnru a discrimina și a înlătura ambiguitatea. Nu manifestă niciun semn că
	Nu manifestă niciun semn că trebuie	trebuie să reducă ceea ce vrea să spună.
	să reducă ceea ce vrea să spună.	irebute su reduca ecca ec vrea su spana.
C1	Poate alege formularea adecvată	Poate alege formularea adecvată dintr-un
	dintr-un repertoriu larg de variante	repertoriu larg de variante pentru a exprima clar,
	pentru a exprima clar, fără restricții,	fără restricții, ceea ce vrea să spună.
	ceea ce vrea să spună.	Posto utiliza adequat si on canceana flouibilitate
		<b>Poate</b> <i>utiliza adecvat</i> și cu <i>oarecare flexibilitate</i> o gamă variată de structuri gramaticale complexe.
B2	Poate să se exprime clar, fără a lăsa	Poate să se exprime clar și fără a lăsa impresia că
	impresia că este nevoit să restrângă	este nevoit să restrângă ceea ce dorește să spună.
	ceea ce dorește să spună.	
	Posedă o gamă destul de bogată de	Posedă o gamă destul de bogată de mijloace
	mijloace lingvistice pentru a reuși să descrie clar, să-și exprime punctul de	lingvistice pentru a reuși să descrie clar, să-și exprime punctul de vedere și să dezvolte o
	vedere și să dezvolte o argumentație	argumentație fără a-și căuta cuvintele în mod
	fără a-și căuta cuvintele în mod	evident, folosind fraze complexe.
	evident, folosind fraze complexe.	-
B1	Posedă o gamă suficient de vastă a	Posedă o gamă suficient de vastă a mijloacelor
	mijloacelor lingvistice pentru a	lingvistice pentru a descrie situații imprevizibile,
	descrie situații imprevizibile, pentru a explica punctul principal al unei	a explica <i>punctul principal al unei probleme</i> sau al unei idei cu destulă precizie, <i>a-şi exprima</i>
	probleme sau al unei idei cu destulă	părerea în legătură cu subiecte abstracte sau
	precizie, pentru a-și exprima părerea	culturale ca muzica sau cinematografia.
	în legătură cu subiecte abstracte sau	
	culturale, cum ar fi muzica sau	
	cinematografia.	D.4' C'
	Deține suficiente mijloace lingvistice pentru a face față, cu	<b>Deține</b> suficiente mijloace lingvistice pentru a face față, cu anumite ezitări și recurgând la
	anumite ezitări și recurgând la	parafrazări, unor subiecte, cum ar fi: familie,
	parafrazări, unor subiecte, cum ar fi:	hobby-uri, timp liber, teme de interes, călătorii și
	familia, hobby-uri, timp liber, teme	teme de actualitate, dar vocabularul limitat
	de interes, călătorii și teme de	conduce la <i>repetiții</i> și uneori chiar la <i>dificultăți de</i>
	actualitate, dar vocabularul limitat	formulare.

<sup>&</sup>lt;sup>114</sup> Companion Volume, p. 131.

<sup>&</sup>lt;sup>115</sup> Ibidem.

	conduce la repetări și uneori chiar la dificultăți de formulare.	
A2	Are un repertoriu elementar de structuri lingvistice care îi permite să facă față situațiilor curente, cu conținut previzibil, deși are în general nevoie să își caute cuvintele și să facă un compromis în raport cu intențiile sale de comunicare.	Are un repertoriu elementar de structuri lingvistice care îi permite să facă față situațiilor curente, cu conținut previzibil, deși are în general nevoie să își caute cuvintele și să facă un compromis în raport cu intențiile sale de comunicare.
	Poate produce scurte expresii uzuale pentru a-și satisface nevoile de tip concret: detalii personale, rutină zilnică, dorințe și nevoi, solicitări de informații.	Poate produce scurte expresii uzuale pentru a-și satisface nevoile de tip concret: detalii personale, rutină zilnică, dorințe și nevoi, solicitări de informații.
	Poate utiliza modele de fraze elementare și poate comunica prin intermediul unor expresii memorate, a unor grupuri de câteva cuvinte și cu ajutorul unor formule fixe despre sine, alți oameni, ceea ce fac, locuri sau bunurile lor.  Deține un repertoriu limitat de scurte expresii memorate care acoperă situații predictibile de supraviețuire; apar frecvent pauze și neînțelegeri în situații neprevăzute.	Poate utiliza modele de fraze elementare și poate comunica prin intermediul unor expresii memorate, a unor grupuri de câteva cuvinte și cu ajutorul unor formule fixe despre sine, alți oameni, ceea ce fac, locuri sau bunurile lor.  Deține un repertoriu limitat de scurte expresii memorate care acoperă situații predictibile de supraviețuire; apar frecvent pauze și neînțelegeri în situații neprevăzute.
A1	Deține un repertoriu de bază de expresii simple în legătură cu detaliile personale și cu nevoile concrete imediate.	Deține un repertoriu de bază de expresii simple în legătură cu detaliile personale și cu nevoile concrete imediate.  Poate utiliza unele structuri de bază în propoziții
		simple, cu unele omisiuni sau reducând unele elemente.
PreA1	nu există	Poate folosi <i>cuvinte izolate</i> și <i>expresii de bază</i> pentru a oferi <i>informații simple despre sine</i> .

Dacă ne referim la scala lingvistică generală, putem remarca, în primul rând, apariția descriptorului pentru nivelul preA1, unde un vorbitor va putea utiliza cuvinte izolate pentru a da informații despre sine. Așadar, în acest caz, vorbim, cel mai probabil, despre prezentarea acestuia, de pildă. În vreme ce, la A1, pe lângă descriptorul păstrat din CECRL, care descria limba *deținută* de către un vorbitor la acest nivel (repertoriu de bază despre sine și nevoi cotidiene imediate), apare unul nou care arată ce poate produce un locutor la acest nivel. Considerăm oportună introducerea acestui descriptor, întrucât experiența ne-a dovedit că, și la acest nivel, putem vorbi totuși despre producții ale nonnativilor. Mai mult, existența sa poate avea si un efect motivant asupra acelora care studiază o limbă străină si au nevoie de

dovezi ale progresului lor. Demn de remarcat, odată cu nivelul A2, este că vorbitorul începe să "comunice", să participe la interacțiuni, reuşind să-și folosească repertoriul limitat de cuvinte și expresii memorate pentru a se face înțeles și a vorbi despre alții, nu doar despre sine.

Cât despre volumul vocabularului<sup>116</sup>, acesta se referă la dezvoltarea și la diversitatea elementelor lexicale și a expresiilor utilizate. Conform *Companion Volume*, în general, dezvoltarea vocabularului se realizează prin lecturi extinse<sup>117</sup>. Conceptele operaționalizate în această scală sunt: 1. *cadrul*, de la *situații de tip concret* (A1) la *tranzacții cotidiene în situații comune* (A2), până la *toate situațiile* (la C2) și 2. *tipul structurilor*, de la *un repertoriu elementar* (A1) la unul *extins* (la C2)<sup>118</sup>.

Tabelul 3 - Scala din CECRL pentru volumul vocabularului

Nivel	CECRL	Companion Volume
C2	Poate utiliza cu ușurință un repertoriu extins de elemente lexicale, inclusiv expresii idiomatice și cuvinte colocviale, având conștiința nivelului de conotație semantică.	Poate utiliza cu uşurință un repertoriu extins de elemente lexicale, inclusiv expresii idiomatice și cuvinte colocviale, având conștiința nivelului de conotație semantică.
C1	Stăpânește un repertoriu vast de elemente lexicale, care fac ca pauzele și momentele de căutare a cuvintelor să nu fie observate.	<b>Stăpânește</b> un <i>repertoriu vast de elemente lexicale</i> , care fac ca pauzele și momentele de căutare a cuvintelor să nu fie observate.
	Bună stăpânire a expresiilor idiomatice și a cuvintelor colocviale.	Poate selecta din mai multe opțiuni lexicale, în aproape toate situațiile, prin exploatarea sinonimelor unor cuvinte încă și mai puțin comune.
		Bună stăpânire a expresiilor idiomatice și a cuvintelor colocviale.
B2	Posedă o gamă variată de elemente de vocabular cu privire la domeniul său de interes, precum și la majoritatea temelor generale.	Poate înțelege și utiliza adecvat gama vocabularului tehnic și a expresiilor idiomatice comune, în domeniul său de specializare.  Posedă o gamă variată de elemente de
	Poate reformula pentru a evita repetițiile frecvente, însă lipsa unor cuvinte încă poate cauza ezitări și parafrazări.	vocabular cu privire la domeniul său de interes, precum și la majoritatea temelor generale.

<sup>&</sup>lt;sup>116</sup> Companion Volume, p. 132.

<sup>&</sup>lt;sup>117</sup> Ibidem.

<sup>&</sup>lt;sup>118</sup> Ibidem.

		<b>Poate reformula</b> pentru a <i>evita repetițiile frecvente</i> , însă lipsa unor cuvinte încă poate cauza ezitări și parafrazări.
		<b>Poate produce</b> sintagme adecvate pentru multe <i>cuvinte</i> , <i>în context</i> , în mod sistematic.
		Poate înțelege și folosi o mare parte din lexicul specializat din domeniul său, însă are probleme cu terminologia de specialitate din afara domeniului său.
B1	Deține un bagaj lexical care îi ajunge pentru a se exprima, cu ajutorul unor	<b>Posedă</b> o gamă de elemente lexicale cu referire la <i>teme familiar</i> și <i>situații cotidiene</i> .
	parafraze, cu privire la majoritatea	, ,
	temelor din viața de zi cu zi, precum: familie, hobby-uri și interese, muncă,	<b>Deține</b> un bagaj lexical care îi ajunge pentru a
	călătorii și teme de actualitate.	se <i>exprima, cu ajutorul unor parafraze,</i> cu privire la majoritatea temelor din viața de zi cu
	calatom și teme de actualitate.	zi, precum: familie, hobby-uri și interese,
		muncă, călătorii și teme de actualitate.
A2	Posedă un vocabular suficient de extins	Posedă un vocabular suficient de extins pentru
	pentru a desfășura activități obișnuite,	a desfășura <i>activități obișnuite</i> , <i>tranzacții</i>
	tranzacții cotidiene cu privire la teme și situații familiare.	cotidiene cu privire la teme și situații familiare.
	Deține un vocabular suficient de extins pentru a putea exprima nevoile comunicative de bază.	<b>Deține</b> un vocabular suficient de extins pentru a putea <i>exprima nevoile comunicative de bază</i> .
		Deține un vocabular suficient de extins pentru
	Deține un vocabular suficient de extins	a putea <i>face față nevoilor simple de</i>
	pentru a putea face față nevoilor simple	supraviețuire.
A1	de supraviețuire.  Deține un repertoriu elementar de	Datina un ranantaniu alamantan da avvieta si
AI	cuvinte izolate și expresii simple în	<b>Deține</b> un repertoriu elementar de cuvinte și expresii simple, în relație cu situații de tip
	relație cu situații de tip curent.	curent.
preA1	Nu există	Nu există descriptor

Ceea ce putem observa, la o primă vedere, la scala pentru volumul vocabularului este că, la primele niveluri, nu apar noi descriptori (abia la B1 apare o nouă descriere). De asemenea, nu există disponibil un descriptor pentru preA1, iar la A1 s-a renunțat la sintagma *cuvinte izolate* în favoarea formulării *cuvinte și expresii simple*, cel mai probabil, deoarece *cuvinte izolate* de acum înainte caracterizează nivelul preA1.

Conform celor două documente lingvistice europene în discuție, stăpânirea vocabularului se referă la "abilitatea utilizatorului, a cursantului, de a selecta, din repertoriul său de structuri lexico-gramaticale, cuvântul, expresia adecvat(ă) situației de comunicare în care se găsește. Caracteristic acestei abilități este faptul că, odată ce e dezvoltată, selecția are loc pe baza asocierilor în forma expresiilor idiomatice și

a colocațiilor<sup>119</sup>, recunoașterea unei expresii declașând accesarea alteia și tot așa" [tr. n.]<sup>120</sup>. Conceptele operaționalizate prin această scală sunt: 1. *familiaritatea temelor* (A1-B1) și 2. *gradul de control* (B2-C2).

Tabelul 4 - Scala din CECRL pentru stăpânirea vocabularului

Nivel	CECRL	Companion Volume
C2	Utilizarea corectă și adecvată a	Utilizarea corectă și adecvată a
	vocabularului, în mod constant.	vocabularului, în mod constant.
C1	Ocazional, apar mici scăpări, dar nu	Utilizează adecvat și în expresii idiomatice un
	erori semnificative de vocabular	vocabular mai puțin comun.
		Ocazional, apar mici scăpări, dar nu erori
		semnificative de vocabular
B2	Acuratețea vocabularului este, în	Acuratețea vocabularului este, în general, la
	general, la nivel înalt. Totuși, pot	nivel înalt. Totuși, pot apărea unele confuzii
	apărea unele confuzii sau alegerea unor	sau alegerea unor cuvinte incorecte, însă fără
	cuvinte incorecte, însă fără a afecta	a afecta comunicarea.
	comunicarea.	
B1	Dovedește o bună stăpânire a	Dovedește o bună stăpânire a vocabularului
	vocabularului elementar, deși se mai	elementar, deși se mai produc erori serioase,
	produc erori serioase, când exprimă	când exprimă idei complexe sau când are de-a
	idei complexe sau când are de-a face cu	face cu teme și situații neprevăzute.
	teme și situații neprevăzute.	Hilimana adanyat a same yaniate da
		Utilizează adecvat o gamă variată de
		elemente simple de vocabular atunci când
A2	Dogto stănâni ya nanontoniy limitat în	abordează teme familiare.
A2	Poate stăpâni un repertoriu limitat în legătură cu nevoile uzuale concrete.	Poate stăpâni un repertoriu limitat în legătură cu nevoile uzuale concrete.
A1	Nu există descriptor	-
		Nu există descriptor
PreA1	Nu există	Nu există descriptor

Analizând scala pentru stăpânirea vocabularului<sup>121</sup>, remarcăm că nu putem vorbi despre controlul lexicului mai devreme de A2, dar și în acest caz există doar un repertoriu limitat de elemente lexicale care pot fi stăpânite, în special cele legate de nevoile concrete.

#### 1.5. Concluzii

Prin urmare, CECRL are meritul de a reprezenta documentul lingvistic european cu cea mai mare influență, la această oră, în ceea ce privește politicile lingvistice europene, în special. În ciuda criticilor aduse de-a lungul vremii, odată cu

<sup>&</sup>lt;sup>119</sup> Prin colocații se înțelege "coocurența sistematică a anumitor cuvinte" (vezi, în acest sens, Lavinia Iunia-Vasiu, *op. cit.*, 2020, p. 104.

<sup>&</sup>lt;sup>120</sup> Companion Volume, p. 134.

<sup>&</sup>lt;sup>121</sup> Ibidem.

apariția sa, s-a făcut un uriaș pas înainte, oferindu-le oamenilor care îl cunosc codul prin care acestia se pot înțelege în plan profesional, indiferent de cultura pedagogică sau de interesele fiecăruia. Mai mult, acesta a stat la baza elaborării multor materiale cu rol fundamental în achiziția limbilor. Recent, odată cu apariția Companion Volume, s-au adus îmbunătățiri considerabile descriptorilor și altor categorii care puteau fi percepute drept ambigue în CECRL. Cu toate acestea, în continuare, considerăm că este o mare nevoie de descriptori și mai clari, și mai specifici, precum și de validarea lor cu ajutorul unor date empirice, bazate pe corpus, pentru a putea pune la dispoziția practicienilor instrumente și mai detaliate în scopul asigurării unui proces de predare-învătare-evaluare cât mai transparent. Considerăm că niste date mai clare, exemple si limitări pot fi obtinute mai ales cu ajutorul analizei corpusului vorbitorilor nonnativi, precum și al inventarului lexical pe nivelurile de referință. Dacă în capitolul acesta am prezentat politica lingvistică europeană asupra L2, în următorul ne concentrăm atenția asupra obiectivului nostru, inventarul lexical al RLS, pentru nivelul A1, și prezentăm care este situația cercetării atât pe plan national, cât si international.

## CAPITOLUL II – OBSERVAȚII ASUPRA STADIULUI ACTUAL AL CERCETĂRII

Importanța covârșitoare a vocabularului în achiziția unei limbi este cunoscută si afirmată deja de multă vreme. Aceasta se poate observa, conform lui Zimmerman, în felul în care alegem cuvinte diferite pentru "a sugera, într-un mod subtil, lucruri diferite" [tr. n.]<sup>122</sup>, adică, așa cum arată autoarea, cuvintele pe care le selectăm, în funcție de situația de comunicare în care ne aflăm, ne oferă control asupra a ceea ce dorim să transmitem<sup>123</sup>. Fiind "conectat direct cu structura gramaticală a limbii, dar si cu contextul extralingvistic în care se integrează, lexicul este un domeniu fluctuant și complex și deci greu de schematizat"<sup>124</sup>. Probabil acesta este si motivul pentru care numărul studiilor destinate achiziției lexicale în domeniul RLS e semnificativ mai mic raportat la aspectele gramaticale, de exemplu. "Achiziția lexicală este adeseori considerată ca automată, asociată superfluu aproprierii structurilor gramaticale sau intrinsec inclusă în procesul de asimilare comunicatională"<sup>125</sup>. Așadar, prin cercetarea noastră ne-am propus să sensibilizăm specialiștii în legătură cu necesitatea de a acorda o mai mare atenție vocabularului, abordând o temă actuală, puțin tratată de către cercetătorii români din domeniul RLS, însă care constituie unul dintre subiectele de mare interes pentru lingviștii europeni, în special, pentru cei anglofoni: realizarea unui inventar lexical, pentru nivelul A1 din CECRL, bazat pe un corpus al nonnativilor. Consultând literatura străină de specialitate, constatăm că, spre exemplu, în cazul limbilor engleză<sup>126</sup>, spaniolă<sup>127</sup> si germană<sup>128</sup>, dar nu numai, au fost deja elaborate inventare lexicale pe niveluri de

<sup>&</sup>lt;sup>122</sup> Cheryl Boyd Zimmerman, *Word Knowledge. A Vocabulary Teacher's Handbook*, New York, Oxford University Press, 2009, p. 16.

<sup>&</sup>lt;sup>123</sup> Cf. Cheryl Boyd Zimmerman, op. cit., 2009, p. 16.

<sup>&</sup>lt;sup>124</sup> Gabriela Stoica, "Dificultăți lexico-semantice în achiziția limbii române ca limbă străină", în Elena Platon, Antonela Arieșan (eds.), *Noi perspective în abordarea românei ca limbă străină/ca limbă nematernă*, Cluj Napoca, Editura Casa Cărții de Știință, 2012, p. 182.
<sup>125</sup> Ibidem.

<sup>&</sup>lt;sup>126</sup> În cazul limbii engleze, avem de-a face cu cel mai extins și complex inventar lexical, mereu în dezvoltare, care se ocupă atât de inventarierea, plasarea și descrierea elementelor lexicale din engleza britanică, cât și de cele din engleza americană, *English Vocabulary Profile* (EVP), elaborat de Cambridge University Press.

<sup>&</sup>lt;sup>127</sup> Ne referim la *El plan curricular del Instituto Cervantes*.

<sup>&</sup>lt;sup>128</sup> În ceea ce privește limba germană, Institutul Goethe a elaborat inventare lexicale pe niveluri de competență lingvistică, pornind de la materialele elaborate în cadrul Institutului în vederea evaluării cunoștințelor de limbă germană. De asemenea, și în cazul limbii germane s-a realizat și un profil

competență lingvistică, dezvoltate în cadrul unor proiecte de mai mici sau mai mari dimensiuni. Înainte de a prezenta situația limbii române ca limbă străină, prezentăm aceste modele, subliniind metoda de cercetare, avantajele si limitele lor.

#### 2.1. Situația altor limbi de circulație internațională

#### 2.1.1. Cazul limbii engleze

După o analiză a materialelor destinate limbii engleze, putem afirma, fără îndoială, că aceasta este limba care dispune de cele mai multe instrumente, existând o mare varietate de liste de cuvinte și inventare, care au fost mereu adaptate și dezvoltate<sup>129</sup>. Pe parcursul ultimului secol, se remarcă nenumărate liste de cuvinte care au influentat modul în care se elaborează acum aceste inventare. Dintre acestea, se remarcă Thorndike, în 1921, care a introdus pentru prima dată criteriul răspândirii (range), și West, în 1953, care a dezvoltat A General Service List of English Words (GSL). Până în prezent, această din urmă lucrare constituie una dintre cele mai importante liste de cuvinte din limba engleză, incluzând "primele 2000 de cuvinte, ca frecvență, precum și frecvențele (relative) ale sensurilor cuvintelor incluse" [tr. n.] 130. De remarcat la această listă este faptul că, pentru elaborarea sa, au fost utilizate atât criterii obiective (frecvența și răspândirea), cât și criterii subiective (de exemplu, usurinta în învătare, necesitatea, suprapunerea, în ceea ce priveste sensul, cu alte cuvinte deja prezente în listă, stil, precum și cuvinte care descriu emoțiile). Odată cu anii 1970, analizele au început să fie făcute electronic, ceea ce a reprezentat o mare reușită și, evident, o ușurare semnificativă a tehnicilor de lucru folosite.

Un alt moment important în acest proces îl constituie *Cadrul comun lexical* pentru limba engleză (CELF), elaborat pe baza Corpusului Rinsland (6 milioane de cuvinte), extras din lucrările elevilor americani de până la clasa a VIII-a<sup>131</sup>. Mai

lingvistic, *Profile Deutsch*, elaborat într-un parteneriat între Germania, Austria și Elveția, la decizia Consiliului Europei.

<sup>&</sup>lt;sup>129</sup> Create în funcție de diferite criterii, cum ar fi: apartenența la o anumită parte de vorbire, frecvența în vorbire/în textele scrise literare sau nonliterare etc. Exemple de astfel de liste se pot găsi la adresa web: http://www.manythings.org/vocabulary/lists/l/

<sup>&</sup>lt;sup>130</sup> I.S.P. Nation, *Making and Using Word Lists for Language Learning and Testing*, John Benjamin Publishing Company, Amsterdam/Philadephia, 2016, p. 10.

<sup>&</sup>lt;sup>131</sup> Diana Burlacu, "Repartizarea lexicului RLNM pe nivelurile CECRL – cu aplicație la ciclul primar", în *Actele Conferinței de lansare a proiectului "Perfecționarea cadrelor didactice din învățământul preuniversitar care predau limba română minorităților naționale*", Cluj-Napoca, Editura Casa Cărții de Stiintă, 2011, p. 234.

târziu s-a ajuns la materialele elaborate de Cambridge University Press: *English Profile Wordlists*, care s-au transformat în *English Vocabulary Profile (EVP)*.

#### English Vocabulary Profile (EVP)

EVP constituie un material online extrem de valoros, care oferă informații atât despre nivelul unui cuvânt, cât și despre sensul acestuia și despre rolul său în alte expresii. EVP este disponibil atât în engleza americană, cât și în cea britanică. Scopul acestui material, după cum apare în descrierea proiectului, este "acela de a oferi o sursă pentru elaborarea curriculumului sau pentru autorii de materiale, teste, profesori sau formatori" [tr. n.]<sup>132</sup>.

#### Corpus și criterii de selecție

EVP este rezultatul comparării corpusului CLC (Cambridge Learner Corpus) – o colecție de sute de mii de producții scrise ale studenților care au învățat limba engleză, din diferite zone și din diferiți ani – cu alte materiale care dovedesc maniera de utilizare a limbii engleze, precum: teste de vocabular, materiale produse în clasă sau suporturi de curs<sup>133</sup>. În plus, pentru ca rezultatele să fie cât mai aproape de realitate, s-a folosit și corpusul aparținând Universității Cambridge, un corpus vast de limbă scrisă și vorbită. În ceea ce privește nivelurile superioare, s-a recurs la materialele destinate vocabularului academic și la dicționarele de frecvență pentru expresii idiomatice. Pe lângă aceste mijloace, variantele finale sunt rezultatul validării listelor din partea unor experți, din afara proiectului, dar și de către unii utilizatori ai materialului (autori de manuale, profesori și, chiar, studenți).

Ceea ce diferențiază EVP de alte studii de acest tip, așadar, este tocmai utilizarea diferitelor corpusuri pentru a susține cât mai convingător cu putință cercetarea.

"Descrierile precedente ale limbilor au fost realizate de către specialiști care și-au folosit, în mare măsură, experiența de vorbitori și de profesori. Prin contrast, folosind analiza corpusului, metodologia EVP este una empirică, oferind

<sup>132</sup> http://www.englishprofile.org/index.php/wordlists, consultat în 07.09. 2014, ora 13:09.

<sup>&</sup>lt;sup>133</sup>Aici este vorba despre listele din manualele cele mai utilizate și din caietele de exerciții de vocabular, listele de cuvinte de la examenele KET și PET (pentru copii), precum și *Cambridge English Lexicon* al lui Roland Hindmarsh, publicat acum 30 de ani, folosit pentru confirmarea unor decizii de plasare pe un anumit nivel.

dovezi clare în ceea ce privește cunoștințele de limbă pe care le au vorbitorii din diferite colțuri ale lumii, la fiecare nivel de competență lingvistică" [tr. n.]<sup>134</sup>.

## Essential Word List, Dang și Webb

Pe lângă modelul EVP, am ales să prezentăm și materialul elaborat de către Dang și Webb, autorii Essential Word List (EWL)<sup>135</sup>, o listă de cuvinte destinată începătorilor de limbă engleză, întrucât aceasta reprezintă un alt mod de realizare a unui instrument lexical. Materialul în discutie este elaborat diferit de ceea ce am prezentat până acum. EWL a fost creată pe baza a patru liste pre-existente<sup>136</sup>. urmărind să răspundă unui set de 7 întrebări și parcurgând mai multe etape<sup>137</sup>. Lista initială a fost creată prin excluderea suprapunerilor datorate combinării celor 4 instrumente folosite, iar, apoi, prin aplicarea criteriilor stabilite. În ceea ce privește determinarea limitei listei, adică identificarea unui scor cut-off sub care nu au mai fost luate în calcul cuvintele, autorii au folosit trei criterii: aplicabilitate (practicability), aparitia unei modificări în curba gradului de acoperire a lexicului (change in the lexical coverage curve) și acoperirea propriu-zisă a lexicului (amount of lexical coverage)<sup>138</sup>. Prin aplicabilitate se întelege, în acest caz, relationarea dimensiunii listei la cantitatea de cuvinte pe care un student de L2 o poate învăta în cadrul unui program de studiu, asadar, obtinerea unei liste într-atât de lungi încât să fie posibil de însușit într-un program de studiu<sup>139</sup>.

Așa cum arată Dang și Webb, pentru a identifica eventualele modificări în curba gradului de acoperire a lexicului a fost examinată fiecare modificare apărută la fiecare 100 de cuvinte avute în vedere, iar prin acoperirea propriu-zisă a lexicului s-a urmărit ce se poate obține cu acele cuvinte introduse în listă, de exemplu, ce procent dintr-un text poate fi acoperit și înțeles<sup>140</sup>. Autorii s-au limitat la 800 de cuvinte, 624 de cuvinte cu conținut lexical (*lexical words*) și 176 de cuvinte de

<sup>&</sup>lt;sup>134</sup> Harrison, Julia, Barker, Fiona (eds.), *English Profile Studies. English Profile in Practice*, Cambridge University Press, Cambridge, 2015, p. 4.

<sup>&</sup>lt;sup>135</sup> Demersul care a stat la baza EWL este descris în Thi Ngoc Yen Dang, Stuart Webb, "Making an essential word list for beginners", în I.S.P. Nation, *op. cit*, 2016, pp. 153-167.

<sup>&</sup>lt;sup>136</sup> General Service List, elaborată de West în anul 1953, BNC2000, din 2006, și BNC/COCA 2000, din 2012, ambele realizate de Nation, și *New-GSL*, elaborată de Brezina și Gablasova în 2013, ultimele trei conținând cele mai recente statistici ale frecvenței cuvintelor în limba engleză.

<sup>&</sup>lt;sup>137</sup> Prezentate în Thi Ngoc Yen Dang, Stuart Webb, op. cit., 2016, p. 155.

<sup>&</sup>lt;sup>138</sup> Thi Ngoc Yen Dang, Stuart Webb, op. cit., 2016, p. 159.

<sup>&</sup>lt;sup>139</sup> Ibidem.

<sup>&</sup>lt;sup>140</sup> Ibidem.

legătură (*function words*)<sup>141</sup>. Această separare nu a mai fost făcută în cazul niciunei alte liste de cuvinte, ceea ce o face mult mai potrivită pentru scopuri pedagogice<sup>142</sup>.

"Într-un text, cuvintele cu conținut lexical au o importanță mult mai mare decât cuvintele de legătură, de aceea e mai important ca aceste cuvinte să fie primele organizate în funcție de frecvența lor în limbă. Așadar, felul în care se abordează cuvintele cu conținut lexical ar trebui să fie diferit de modul în care se abordează cuvintele de legătură" [tr. n.]<sup>143</sup>.

Autorii au fost nevoiţi să ia unele decizii cu privire la ce trebuie inclus în EWL. Astfel, au hotărât să nu includă în inventarul lor literele alfabetului, întrucât au presupus că studenții care învață o limbă străină ar trebui să cunoască deja alfabetul, tocmai datorită importanței sale cruciale în însușirea unei noi limbi<sup>144</sup>. Nici afixele nu au fost luate în considerare, deși acestea au o importanță mare pentru studenții de L2, argumentul principal în luarea acestei decizii constând în faptul că grupul-țintă s-ar putea să nu aibă suficiente cunoștințe gramaticale pentru a și le însuși, având în vedere că vorbim despre studenții începători<sup>145</sup>.

Un alt aspect care conferă unicitate EWL este reprezentat de fragmentarea acesteia în 13 subliste, organizate descrescător, în funcție de frecvență și de capacitatea de acoperire lexicală. Această împărțire le face mult mai ușor de folosit, întrucât utilizatorul nu va fi "șocat" de o listă interminabilă de cuvinte, pe care ar putea să o gestioneze mult mai greu. Astfel, primele 12 liste conțin fiecare câte 50 de cuvinte, în vreme ce lista a 13-a are doar 24 de cuvinte. Cel mai mare avantaj al unei astfel de segmentări a listei, prezentat de către autori, este acela că, "dacă va fi urmată această listă, va crește eficacitatea predării și a învățării, astfel având siguranța că, prima dată, vor fi prezentate doar cele mai utile elemente lexicale. De asemenea, lista le permite autorilor programelor de studiu să-și organizeze curriculumul" [tr. n.]<sup>146</sup>, evitându-se suprapunerile care în alte circumstanțe sunt inevitabile.

Dacă privim comparativ cele două instrumente destinate limbii engleze, remarcăm două direcții diferite, în funcție de destinatar. În ciuda aceluiași obiectiv enunțat, acela de a constitui un instrument utlizat în proiectarea cursurilor, a

<sup>&</sup>lt;sup>141</sup> Thi Ngoc Yen Dang, Stuart Webb, op. cit., 2016, p. 166.

<sup>&</sup>lt;sup>142</sup> Ibidem.

<sup>&</sup>lt;sup>143</sup> Ibidem.

<sup>&</sup>lt;sup>144</sup> *Ibidem*, p. 156.

<sup>&</sup>lt;sup>145</sup> Ibidem.

<sup>&</sup>lt;sup>146</sup> *Ibidem*, p. 167.

curriculelor și a materialelor de predare/învățare/evaluare, EVP se distinge de EWL prin faptul că poate fi utilizat și de către student, nu doar de către profesor, putând oricând reprezenta un instrument valoros în cazul unui autodidact.

## Programe online pentru identificarea nivelului și simplificarea textelor

Este bine-cunoscut faptul că, "dacă un curs sau un text conține prea multe cuvinte necunoscute pentru cel care învață, atunci acesta va face un efort mult prea mare pentru a-și explica aceste cuvinte, astfel încât va pierde esența textului" [tr. n.]<sup>147</sup>. De exemplu, un text devine prea dificil de înțeles pentru oricine dacă pe fiecare rând se regăsesc unul sau două cuvinte necunoscute. Acesta este motivul pentru care s-au dezvoltat programe care, pe baza unor inventare lexicale bine elaborate, pot identifica elementele care sunt susceptibile să facă înțelegerea dificilă și pot chiar propune soluții de simplificare.

Primul instrument lingvistic online la care facem referire este *Text inspector*, care are drept rol principal analiza vocabularului conținut în diverse texte, în special în achiziția L2. Așa cum se declară chiar pe pagina oficială a acestui instrument lingvistic, "Text Inspector analizează dificultatea textelor în limba engleză, oferindui utilizatorului, imediat ce introduce un material în acest program, un scor, precum și un feedback detaliat pe textul introdus"<sup>148</sup> astfel încât utilizatorul să își poată modifica textul pentru a putea fi folosit în scopurile și la nivelul dorite. Pentru a veni în sprijinul cercetătorilor și a profesorilor de L2, acest material a fost adaptat, putând fi folosit în multe alte limbi<sup>149</sup>, după cum a declarat chiar creatorul său, Stephen Bax, în cadrul unui workshop dedicat *Text Inspector*, în anul 2017, la Bologna, la conferința ALTE, *Learning and Assessment: Making the Connections*, atelier la care am luat parte și noi.

Un alt exemplu este reprezentat de site-ul web *Compleat Word Tutor*<sup>150</sup>, dezvoltat de către Tom Cobb pentru a ilustra ideea conform căreia, "dacă un cuvânt e întâlnit în contexte variate, atunci acesta va fi însușit mult mai ușor și va fi folosit într-un mod mult mai flexibil" [tr.n.]<sup>151</sup>. Site-ul este axat atât pe engleză, cât și pe spaniolă si franceză, constituind un valoros instrument pentru cei care se

<sup>&</sup>lt;sup>147</sup> I.S.P. Nation, op. cit., 2016, p. 5.

<sup>148</sup> https://textinspector.com.

<sup>149</sup> Totuși, în momentul în care a avut loc acest atelier de lucru, Stephen Bax a evidențiat unele dificultăți întâmpinate în cazul limbilor care folosesc semne diacritice, ceea ce a făcut ca, pentru aceste limbi, să nu fie încă disponibil materialul.

<sup>&</sup>lt;sup>150</sup> Disponibil la adresa web: www.lextutor.ca.

<sup>&</sup>lt;sup>151</sup> Marti Sevier, "The Compleat Tutor, v.4", in *Teaching English as a Second or Foreign Language*, vol. 8, no. 3, december 2004.

concentrează asupra vocabularului utilizat în scopuri academice. În cadrul paginii destinate cercetării, Tom Cobb explică rațiunea pentru care studenții folosesc liste de cuvinte: "dacă știi primele 2000 de cuvinte dintre cele mai frecvent folosite în limba engleză, plus, cele 570 de cuvinte din lista limbajului academic, atunci cunoști aproximativ 90% dintre cuvintele pe care le vei întâlni în orice text academic" [tr.n.]<sup>152</sup>. Pe lângă rolul său în identificarea nivelului și a dificultății textelor, *Compleat Word Tutor* constituie o resursă și în ceea ce privește elaborarea exercițiilor pe baza textelor analizate. Aplicații similare sunt și *AntWordProfiler*<sup>153</sup>, respectiv *Range*<sup>154</sup>.

După elaborarea listelor de cuvinte pe niveluri, considerăm că ar fi extrem de util și pentru domeniul RLS un astfel de program informatic, întrucât ar ajuta foarte mult profesorii în adaptarea pe nivel a textelor folosite la cursuri. Printr-un astfel de program informatic, s-ar putea controla mult mai bine cuvintele predate și n-ar fi neglijate unele în defavoarea altora, întrucât se știe că învățarea se bazează pe repetiție, adică reîntâlnirea cu anumite cuvinte. În plus, și utilizarea acestor elemente lexicale în contexte variate va asigura o însușire mult mai eficientă a acestora<sup>155</sup>.

## 2.1.2. Profile Deutsch și inventarul lexical al limbii germane

În ceea ce privește limba germană, la Institutul Goethe s-au realizat inventare lexicale pe niveluri de competență lingvistică, pornind de la materialele elaborate în cadrul instituției în vederea pregătirii pentru testele de germană<sup>156</sup>. Observăm în cazul acestui material un alt demers în elaborare. Numeroasele materiale disponibile și folosite, realizate plecând de la ceea ce poate face un vorbitor la fiecare nivel, fac ca și această metodă să fie una validă. De asemenea, și în cazul limbii germane (ca în cazul limbii engleze și, după cum vom vedea, și al limbii spaniole<sup>157</sup>) s-a realizat un profil lingvistic, *Profile Deutsch*, elaborat într-un parteneriat între Germania, Austria și Elveția, la decizia Consiliului Europei. Conform autorilor, scopul acestui material a fost acela de a oferi un suport în evaluarea pe nivel și stabilirea unor obiective educaționale în planificarea, predarea

<sup>153</sup> Anthony, L. (2014). *AntWordProfiler* (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software.

<sup>152</sup> https://www.lextutor.ca/research/.

<sup>&</sup>lt;sup>154</sup>Disponibil online, pentru limba engleză, sub egida Universității Victoria, Wellington, Noua Zeelandă: https://www.wgtn.ac.nz/lals/resources/vocabulary-analysis-programs.

<sup>&</sup>lt;sup>155</sup> I.S.P. Nation, op. cit., 2016, p. 5.

<sup>&</sup>lt;sup>156</sup> Peter Kiefer (coord.), *Goethe-Zertifikat A1, Start Deutsch 1*, Prüfungsziele Testbeschreibung, Goethe Institut, 2011.

<sup>&</sup>lt;sup>157</sup> Precum și în cazul altor limbi pe care nu le descriem aici, însă despre care se pot afla mai multe informații la adresa web: http://www.coe.int/t/dg4/linguistic/DNR EN.asp#P61 9444.

și evaluarea unor lecții, precum și în elaborarea materialelor destinate predării și evaluării<sup>158</sup>.

#### 2.1.3. Situația limbii spaniole

După materialele realizate pentru engleză și germană, deja prezentate mai sus, trebuie spus că și spaniola dispune de un inventar lexical pe niveluri, în cadrul unui impresionant material exhaustiv, *El plan curricular del Instituto Cervantes*, în care Institutul Cervantes și-a propus să descrie și să stabilească nivelurile de competență, conform CECRL. În plus,

"acest *Plan Curricular* prezintă tratarea sistematică a materialelor necesare pregătirii obiectivelor și a conținuturilor programelor de la Institutul Cervantes. În același timp, materialul a fost dezvoltat cu ideea de a oferi un instrument util pentru profesorii de limba spaniolă din întreaga lume. Consolidarea calității predării în rețeaua de institute Cervantes constituie obiectivul imediat al acestei publicații, îndeplinind, totodată, obiectivul general, acela de *a contribui la dezvoltarea unor programe pentru îmbunătățirea instrumentelor de analiză și descriere a limbii spaniole*" [tr.n.]<sup>159</sup>.

Așadar, în cadrul acestei publicații, disponibilă și online<sup>160</sup>, sunt descrise toate aspectele limbii spaniole, pe niveluri de competență: gramatica, pronunția, ortografia, funcțiile comunicative și actele de vorbire, tipurile de texte, noțiunile generale (sferele lexicale), noțiunile specifice (inventar lexical pe niveluri), referințele culturale, cunoștințele și comportamentele socio-culturale sau procedurile de învățare.

Spre deosebire de materialul destinat limbii engleze, care este destinat atât vorbitorilor de engleză americană, cât și britanică, în acest caz s-au folosit cuvintele din spaniola vorbită în Spania, și de aceea utilizatorii inventarului trebuie să aibă în vedere și specificul regional al limbii lor, modificând lista în funcție de nevoile de învățare ale cursanților. În acest sens, se oferă un exemplu:

"Seria: *plátano, manzana, naranja* din tema *Alimente* este inclusă în inventar la A2, din cauză că este vorba despre niște fructe consumate pe scară largă în Spania și de aceea s-a considerat că trebuie să facă parte din cunoștințele elevului la acest nivel. Însă e de la sine înțeles că, într-o altă situație de învățare, într-un

159 http://cvc.cervantes.es/ensenanza/biblioteca\_ele/plan\_curricular/default.htm.

<sup>158</sup> http://www.osd.at/default.aspx?SIid=23&LAid=2 (Profile Deutsch).

<sup>&</sup>lt;sup>160</sup> La adresa web: http://cvc.cervantes.es/ensenanza/biblioteca\_ele/plan\_curricular/indice.htm.

centru de limbi dintr-o țară din America Latină, unde fructele sunt diferite, se vor introduce mai întâi cele specifice zonei, de exemplu, piersica sau mango" [tr.n.]<sup>161</sup>.

#### Criteriile de selecție

Pornind întotdeauna de la aprecierea intuitivă a nivelului, bazată pe experiența de predare, autorii au folosit în identificarea nivelului criteriile de frecvență și de utilizare (performanță) în comunicare. Astfel, s-a ținut cont de tipul de texte (scrise și orale) cu care cursanții au de-a face la fiecare nivel (din punctul de vedere al receptării și al producerii). De asemenea, s-a avut în vedere și registrul limbii, în cazul nivelurilor superioare, pentru a asigura precizia și adecvarea în comunicare<sup>162</sup>.

#### 2.1.4. Concluzii

Modelele prezentate în cadrul lucrării de față reprezintă doar o mică parte din ceea ce s-a realizat la nivel european în ceea ce privește descrierea limbilor, pe nivelurile de competență lingvistică din *CECRL* (în momentul de față, țări precum Franța, Italia, Grecia, Cehia, Georgia etc. au realizat sau mai realizează astfel de materiale), însă scopul nostru nu a fost descrierea în detaliu a modelelor elaborate pentru alte limbi, ci acela de a arăta preocuparea crescândă a cercetătorilor pentru acest subiect și deplina convingere a acestora în utilitatea unui asemenea demers.

#### 2.2. Situația actuală a cercetării pe teren românesc

Deși pe plan internațional există un interes tot mai mare pentru acest subiect (vezi *supra* 2.1.1-2.1.3), acesta nu a beneficiat de o atenție deosebită din partea cercetătorilor români. Chiar și așa, de-a lungul timpului, s-au realizat o serie de liste de cuvinte, care au avut ca grup-țintă atât publicul larg, cât și studenții străini. Chiar dacă cele mai multe nu sunt de dată recentă, valoarea lor nu poate fi pusă la îndoială, ele constituind pași importanți și absolut necesari în evoluția viziunii asupra bagajului lexical aferent fiecărui nivel. Prezentăm, în continuare, materialele reprezentative din punctul nostru de vedere.

162 http://cvc.cervantes.es/ensenanza/biblioteca\_ele/plan\_curricular/default.htm.

<sup>&</sup>lt;sup>161</sup> http://cvc.cervantes.es/ensenanza/biblioteca ele/plan curricular/default.htm.

#### 2.2.1. Dictionarul de frecvență (Frequency dictionary of Romanian Words)

Primul material prezentat, *Dicționarul de frecvență al limbii române*<sup>163</sup>, reprezintă un instrument esențial atât din punct de vedere didactic, cât și lexicografic, făcând parte, așa cum se arată în *Cuvântul-înainte* al materialului, dintr-o colecție de dicționare dedicate studierii structurii limbilor romanice<sup>164</sup>. Analiza a avut la bază "o mostră reprezentativă de română modernă, constând din mai mult de 20.000 de propoziții, care au totalizat aproximativ 500.000 de cuvinte" [tr.n.]<sup>165</sup>, colectată din literatura din perioada 1890 - 1940<sup>166</sup>. Conform autorilor<sup>167</sup>, corpusul (*the universe*<sup>168</sup>) folosit pentru obținerea dicționarului, conținând mostre de limbă din cinci genuri diferite, a fost separat în 5 subcorpusuri (*worlds*<sup>169</sup>) egale: literatură dramatică (exclusiv, piese de teatru), ficțiune (nuvele și povestiri), eseistică (eseuri, memorii și corespondență), periodice (editoriale, știri, publicitate) și texte tehnice (medicină, botanică, inginerie, agronomie etc.).

Criteriile folosite în analiza corpusului au fost: frecvența, dispersia și criteriul uzajului. Autorii arată necesitatea utilizării unui alt criteriu decât cel al frecvenței atunci când rezultatele analizei statistice "nu corespund așteptărilor *intuitive* formulate de către specialist bazându-se pe alte informații" [tr.n.]<sup>170</sup>. Astfel s-a ajuns la concluzia că la fel de necesare sunt și criteriul dispersiei și cel al uzajului, care presupune folosirea "dispersiei ca metodă de corecție a frecvenței și viceversa" [tr.n.]<sup>171</sup>.

În urma analizei, *Dicționarul de frecvență* conținea 4.843 de cuvinte, cărora autorii le-au adăugat încă 157 de cuvinte care s-au aflat inițial sub limita inferioară de 3.08, stabilită din perspectiva criteriului uzajului. S-a făcut această modificare pentru ca autorii să poată calcula mai ușor diferitele procente necesare, raportându-se, în acest fel, la 10 clase a câte 500 de cuvinte fiecare, prezentate în partea a doua a dicționarului, sub forma a trei coloane, fiecare corespunzând analizei din perspectiva celor trei criterii<sup>172</sup>. Așadar, conform autorilor această valoare (5000) reprezenta, la momentul realizării materialului, vocabularul limbii române de bază<sup>173</sup>. Probabil că,

<sup>&</sup>lt;sup>163</sup> A. Juillard, P.M.H. Edwards, Ileana Julliard, Frequency dictionary of Romanian Words, Mouton & Co, The Hague, 1965.

<sup>&</sup>lt;sup>164</sup> A. Juillard, P.M.H. Edwards, Ileana Julliard, op. cit., 1965, p. V.

<sup>&</sup>lt;sup>165</sup> *Ibidem*, p. XI.

<sup>&</sup>lt;sup>166</sup> *Ibidem*, p. XIV.

<sup>&</sup>lt;sup>167</sup> Ibidem, p. XV.

<sup>&</sup>lt;sup>168</sup> Termenul utilizat de către autori pentru a desemna corpusul.

<sup>&</sup>lt;sup>169</sup> Termenul utilizat de către autori pentru a numi subcorpusurile.

<sup>&</sup>lt;sup>170</sup> A. Juillard, P.M.H. Edwards, Ileana Julliard, op. cit., 1965, p. XL.

<sup>&</sup>lt;sup>171</sup> *Ibidem*, p. XLVII.

<sup>&</sup>lt;sup>172</sup> Cf. A. Juillard, P.M.H. Edwards, Ileana Julliard, op. cit., 1965, p. XLIX.

<sup>&</sup>lt;sup>173</sup> Ibidem.

în prezent, unele dintre intrările acestui dicționar nu mai pot fi considerate actuale, însă, cu toate acestea, valoarea acestui material lexicografic pentru limba română este covârsitoare.

Observând intrările dicționarului, printre primele cuvinte înregistrate se disting cuvinte de legătură (de, și, în, să, la, nu, pe, mai etc.), verbe simple (a fi, a avea, a face, a trebui etc.) și unele pronume (se, el, care, ei, ce etc.). Totuși, remarcăm, printre primele 200 de intrări, cuvinte care, din punctul nostru de vedere, nu caracterizează nivelul A1, precum: urmă, viață, a părea, suflet, putere etc. Desigur că acest dicționar nu este un instrument elaborat în special pentru a fi utilizat în domeniul RLS. Cu toate acestea, rangul cuvintelor inventariate poate constitui un bun indicator al importanței anumitor cuvinte pentru comunicarea în limba română, în general, așadar, probabil având un rol semnificativ în actul predării românei vorbitorilor nonnativi. Mai mult, amintim că acest volum a reprezentat punctul de plecare în analiză și pentru alte materiale, precum Vocabularul minimal al limbii române pentru studenți străini sau Vocabularul reprezentativ al limbilor romanice (vezi infra 2.2.3 și 2.2.10).

## 2.2.2. Dicționar practic pentru studenți străini. 1000 de cuvinte românești

Dintre materialele destinate studenților, considerăm că *Dicționarul practic* pentru studenți străini. 1000 de cuvinte românești<sup>174</sup>, de Păltineanu, apărut la Universitatea Babeș-Bolyai, în anul 1980, reprezintă un instrument cu relevanță deosebită pentru tema cercetării noastre. Acest dicționar este adresat "studenților străini, în prima etapă a învățării limbii române. Pentru aceasta, se dă un număr de cuvinte, dintre cele mai uzuale, cu sumare indicații gramaticale, știut fiind că un fond de circa o mie de cuvinte, plus gramatica elementară sunt suficiente pentru o conversație simplă într-o limbă străină"<sup>175</sup>. Dacă e să judecăm acest material prin raportare la ceea ce se elaborase deja pe plan internațional, trebuie spus că această publicație apare destul de târziu, când, de pildă, limba engleză beneficia deja de cinci ani de o descriere a nivelului prag, *The Treshold Level*<sup>176</sup>. Faptul că a fost elaborat pentru uzul studenților străini și că organizarea intrărilor a fost realizată în funcție de momentul la care se considera că ar trebui predate ne indică preocuparea autorului pentru ideea de progresie și nevoia acestuia de a stabili o minimă formă de

<sup>176</sup> J.A. van Ek, *The Treshold Level*, Strasbourg, Council of Europe, 1975.

<sup>&</sup>lt;sup>174</sup> Viorel Păltineanu, Dicționar practic pentru studenți străini. 1000 de cuvinte românești, Universitatea Babeș-Bolyai, Cluj-Napoca, 1980.

<sup>&</sup>lt;sup>175</sup> *Ibidem*, p. I.

ierarhizare/repartiție a cuvintelor pe diverse paliere de învățare a limbii, la o oră la care nu se vorbea încă pe teren românesc despre *niveluri*. De remarcat că, dintre cele 395 de cuvinte înregistrate cu majuscule, majoritatea sunt cuvinte de legătură, unele verbe de întrebuințare largă, precum și unele substantive, aspect care asigură o minimă productivitate în limbă, demonstrându-se astfel că, beneficiind de un număr relativ mic de cuvinte "potrivite", se poate comunica.

Autorul afirmă că selecția acestor 1000 de cuvinte s-a realizat "pe baza lucrărilor de statistică lexicală"<sup>177</sup>, a experienței personale și a unui "fond minim de cuvinte vorbite, culese din convorbirile cotidiene"<sup>178</sup>. Desigur că multe dintre elementele lexicale înregistrate în acest document nu mai pot fi considerate de aceeași importanță ca la ora înregistrării. De exemplu, luând la întâmplare elemente lexicale precum *abur, a bănui* sau *tanti*, vom observa, din simpla parcurgere a descriptorilor de nivel A1 și A2 din *CECRL*, că aceste cuvinte, chiar dacă au o importanță mare în limba română, nu pot fi considerate ca elemente lexicale esențiale pentru nivelul A1 și nici chiar pentru nivelul A2<sup>179</sup>, când vorbitorul "posedă un vocabular suficient pentru a satisface *nevoile comunicative elementare*"<sup>180</sup>, iar cuvinte ca cele extrase din materialul în discuție nu mai pot fi considerate, astăzi, niște elemente lexicale absolut necesare într-o comunicare pe teme elementare.

## 2.2.3. Vocabularul minimal al limbii române pentru studenți străini

Dictionarului practic i se adaugă, la doar un an de la apariție, poate cel mai cunoscut material dedicat lexicului RLS, Vocabularul minimal al limbii române pentru studenți străini<sup>181</sup> (pe scurt, VM), coordonat de către Maria Iliescu. Acesta a apărut în trei ediții (1981, 1994, 2001), cu traduceri în mai multe limbi străine (engleză, franceză, spaniolă, germană, italiană). Volumul conține, în ediția din 1981, 3216 "din cele mai importante cuvinte în ce privește frecvența și folosirea în viața zilnică citadină"<sup>182</sup>, în vreme ce în ediția din 1994 au fost incluse 3662 de cuvinte, iar în cea din 2001, 4020. Pe lângă inventarierea elementelor lexicale, acest material oferă și o sistematizare gramaticală în toate cele patru limbi. Selecția cuvintelor s-a

<sup>179</sup> Am ales să argumentăm numai cu nivelurile A1-A2, întrucât aceste 1000 de cuvinte nu pot depăși numeric volumul elementelor lexicale aferente celor două niveluri corespunzătoare CECRL.

<sup>&</sup>lt;sup>177</sup> Viorel Păltineanu, op. cit., 1980, p. II.

<sup>178</sup> Ibidem

<sup>&</sup>lt;sup>180</sup> CECRL, p. 93.

<sup>&</sup>lt;sup>181</sup> Adriana Costăchescu, Maria Iliescu, *Vocabularul minimal al limbii române curente. Cu indicații gramaticale complete*, București, Demiurg, 1995.

<sup>&</sup>lt;sup>182</sup> Maria Iliescu (coord.), *Vocabularul minimal al limbii române pentru studenți străini*, București, Editura Didactică și Pedagogică, 1981, p. 5.

realizat pe baza mai multor criterii, prezentate în continuare: 1. cuvintele din fondul principal lexical al limbii române, 2. cuvinte folosite frecvent în presa anilor 1989, 3. lista cuvintelor din limba administrativă, publicistică, științifică și cea a literaturii, 4. primele 2000 de cuvinte din dicționarul de frecvență<sup>183</sup>, 5. o listă de cuvinte din 14 domenii diferite, obținută pe baza unei anchete în mediul studențesc, și 6. cuvinte introduse după "simț".

Chiar dacă este destinat studenților străini, faptul că nu există o ierarhizare a elementelor înregistrate îngreunează utilizarea acestui material, care "nu poate funcționa drept reper pentru evaluarea cunoștințelor lexicale de RLS și pentru o ierarhizare pe niveluri de studiu, necesitând revizuiri și cercetări extinse"<sup>184</sup>. Mai mult, comparându-l cu alte materiale destinate limbilor de circulație internațională, VM se dovedește a fi mult prea extins, constituind, "cel mai probabil, bagajul lexical al unui vorbitor de RLS cu nivel C2"<sup>185</sup>.

Datorită notorietății sale în domeniul RLS, VM a fost, și încă este, utilizat în elaborarea altor materiale destinate însușirii lexicului RLS. Prezentăm, în continuare, două dintre cele mai cunocute materialele dezvoltate pornind de la VM.

## 2.2.4. Dicționarul minimal al limbii române cu indicații gramaticale uzuale și traducere în franceză, engleză, spaniolă

Dicționarul minimal al limbii române cu indicații gramaticale uzuale și traducere în franceză, engleză, spaniolă (pe scurt, DMLR) constituie, deopotrivă, un material lexicografic, dar și o "gramatică alfabetică"<sup>186</sup>, după cum subliniază autorii în argumentul materialului. DMLR a fost realizat pe baza intuiției și a experienței didactice a celor doi autori, fiind destinat elevilor, studenților români sau străini și profesorilor. Acesta conține un număr de 4000 de cuvinte din româna actuală, selectate din VM, dar și "cuvinte sau îmbinări ne(stabile) din lexicul manualelor de limba română pentru începători"<sup>187</sup>, grație experienței didactice a autorilor. Acestor două categorii li s-au adăugat cuvinte și expresii cu frecvență

<sup>&</sup>lt;sup>183</sup> A. Juillard, P.M.H. Edwards, Ileana Julliard, op. cit., 1965.

<sup>&</sup>lt;sup>184</sup> Gabriela Biriş, "Vocabularul minimal al limbii române: stadiul cunoştinţelor, perspective", în Elena Platon, Antonela Arieşan (eds.), Noi perspective în abordarea românei ca limbă străină/ca limbă nematernă, Cluj Napoca, Editura Casa Cărții de Știință, 2012, p. 180.

<sup>&</sup>lt;sup>185</sup> Gabriela Biriş, *op.cit.*, p.174.

<sup>&</sup>lt;sup>186</sup> Florin-Lazăr Ionilă, Petre Avram, *Dicționar minimal al limbii române cu indicații gramaticale uzuale și traducere în franceză, engleză, spaniolă*, București, Niculescu, 1996, p. 7 (argument).

<sup>&</sup>lt;sup>187</sup> Florin-Lazăr Ionilă, Petre Avram, op. cit., 1996, p. 7 (argument).

ridicată din realitatea social-politică românească de după 1989, precum și termeni religioși sau termeni fundamentali din limbajul lingvistic<sup>188</sup>.

După cum se poate observa și din titlu, lucrarea dispune de traduceri în limbile engleză, franceză și spaniolă, după modelul materialului coordonat de către Iliescu. Deși vorbim despre un instrument util pentru domeniul lexicografic, nu îl putem considera un inventar lexical RLS viabil, pe de-o parte, din pricina vechimii sale, așadar, a inactualității unora dintre termeni, iar, pe de altă parte, deoarece nu este ierarhizat pe niveluri.

# 2.2.5. Antonime. Sinonime. Analogii. Vocabular minimal al limbii române (cu traducere în limba engleză)

Autoarele materialului intitulat Antonime. Sinonime. Analogii. Vocabular minimal al limbii române (cu traducere în limba engleză) sau ASA<sup>189</sup> propun o lucrare lexicografică diferită, în care, pe baza unui număr redus de elemente lexicale, au tratat cuvintele "pornind de la ideea că o unitate lexicală este un ansamblu structurat, un microsistem, cu relații complexe în cursul comunicării – de antonimie, de sinonimie etc. –, si că această unitate lexicală poate fi inclusă în serii analogice, relațiile de sens stabilindu-se contextual"<sup>190</sup>. La întocmirea listei s-au folosit ultimele două ediții ale VM, cele din 1994 și 2001. "Cuvintele selectate sunt majoritatea polisemantice, ceea ce a avut ca rezultat o listă minimală de intrări (676), cu un număr mare de perechi antonimice sau analogii (1417), grupate în 642 de casete, și cu un număr și mai mare de sinonime (peste 2500)"191. O altă trasătură distinctivă a acestui material o reprezintă exemplele, mai accesibile pentru cel care consultă dicționarul, întrucât nu au fost preluate din texte literare, așa cum de cele mai multe ori se întâmplă, ci au fost construite sau adaptate, pe baza experienței autoarelor în predarea și evaluarea RLS. După cum se declară în *Precizările preliminare*, dicționarul se adresează în primul rând vorbitorilor nativi, dar și celor nonnativi. Nu s-a făcut o separare pe nivelurile din CECRL, întrucât, așa cum subliniază și autoarele, "pentru limba română nu există o ierarhizare a lexicului plasat pe cele sase niveluri"<sup>192</sup>, însă e destinat tuturor nivelurilor, datorită prezenței traducerilor în limba

<sup>&</sup>lt;sup>188</sup> *Ibidem*, p. 8 (argument).

<sup>&</sup>lt;sup>189</sup> Gabriela Biriş, Diana Viorela Burlacu, Elisabeta Şoşa, *Antonime. Sinonime. Analogii. Vocabular minimal al limbii române (cu traducere în limba engleză)*, București, Saeculum I.O., 2011.

<sup>&</sup>lt;sup>190</sup> *Ibidem*, p. vi-vii.

<sup>&</sup>lt;sup>191</sup> *Ibidem*, p. viii.

<sup>&</sup>lt;sup>192</sup> *Ibidem*, p. x.

engleză, atât ale cuvintelor, cât și ale exemplelor. Totuși, în ceea ce privește exemplele, "majoritatea se încadrează la nivelurile B1-B2"<sup>193</sup>.

#### 2.2.6. Nivelul Prag pentru învățarea limbii române ca limbă străină

În anul 2001, a apărut ediția în limba română a *The Threshold Level*, intitulată *Nivel Prag pentru învățarea limbii române ca limbă străină*<sup>194</sup>, pentru multă vreme singurul și cel mai complex material care descria cunoștințele și abilitățile maximale pe care un vorbitor ar trebui să le dețină la final de B1. Abordarea "funcțional-noțională", adoptată odată cu introducerea conceptului de *nivel prag*, în anii 1970,

"pune pe prim-plan, dincolo de formele lingvistice folosite în scopul comunicării, ceea ce vrem să facem cu limba pentru a comunica în mod eficient. Desigur, ambele obiective sunt necesare, dar inversarea ordinii de descriere a avut două efecte pozitive. Pe de o parte, s-a ținut cont mai direct de cerințele și motivațiile celor care învață și s-a oferit conceptorilor de metode o imagine foarte detaliată asupra a ceea ce au ei de prezentat. Pe de altă parte, clasificarea funcțiilor și a noțiunilor lingvistice care trebuie exprimate era aplicabilă tuturor limbilor, desigur tuturor celor folosite în spațiul cultural european general" 195.

Abordarea aceasta și-a dovedit eficiența de-a lungul timpului în proiectarea cursurilor și a testelor pentru mai bine de 20 de limbi europene<sup>196</sup>, fapt care a contribuit la îmbrățișarea acesteia și de către specialiștii români prin realizarea *Nivelului Prag*, dorindu-se a fi "un stimul și o bază fermă pentru revitalizarea predării limbii române ca limbă străină sau ca limbă a doua"<sup>197</sup>. Așa cum afirmă și autoarele, "în predarea limbilor străine, opțiunea pentru o astfel de descriere reprezintă reflexul practic al convingerii că studiul gramaticii, al lexicului sau al structurii frazelor nu sunt scopuri în sine, ci mijloace pentru realizarea comunicării"<sup>198</sup>. Astfel, lucrarea conține un inventar de acte de vorbire, cuvinte, structuri lexicale și reguli gramaticale considerate a fi necesare pentru a dobândi

63

<sup>193</sup> Gabriela Biris, Diana Viorela Burlacu, Elisabeta Sosa, op. cit. 2011, p. x.

<sup>&</sup>lt;sup>194</sup> Victoria Moldovan, Liana Pop, Lucia Uricaru, op. cit, 2001.

<sup>&</sup>lt;sup>195</sup> J.L.M. Trim, "Prefață la Nivel Prag", în Victoria Moldovan, Liana Pop, Lucia Uricaru, *op.cit.*, 2001, p. 8.

<sup>&</sup>lt;sup>197</sup> Victoria Moldovan, Liana Pop, Lucia Uricaru, *op.cit.*, 2001, p. 11.

<sup>&</sup>lt;sup>198</sup> Ibidem.

nivelul B1. În elaborarea materialului, "un rol important l-a avut experiența personală a fiecărui autor atât în domeniul predării românei ca limbă străină, cât și al producerii de materiale didactice necesare predării (manuale, gramatici)"<sup>199</sup>. Prin urmare, nu s-au folosit metode empirice sau statistice pentru a se stabili apartența elementelor la categoriile în care au fost incluse. De remarcat că *Nivelul Prag* nu se dorește a fi un cadru prestabilit, folosit de toată lumea, indiferent de grupul-țintă, ci

"lucrarea, în ansamblul ei, trebuie înțeleasă și privită în primul rând ca un instrument metodologic necesar învățării/predării limbii române ca limbă străină și destinat utilizatorilor (organizatori de cursuri, autori de manuale și materiale didactice auxiliare); aceștia își vor selecta pentru a-și construi obiectivele și conținuturile de învățare în funcție de nevoile proprii și de publicul-țintă" 200.

În ceea ce privește lexicul, *noțiunile specifice*, așa cum apar denumite în *Nivelul Prag*, sunt prezentate în jurul a 15 teme<sup>201</sup>. Ceea ce i se poate imputa acestui material e faptul că, fiind elaborat pe baza experienței de predare, nu putem fi în totalitate convinși că propunerea de conținuturi pentru fiecare categorie este și cea potrivită, mai ales că nu e precedată de o astfel de descriere și pentru nivelurile inferioare (A1 și A2). Cu toate acestea, valoarea materialului rămâne de necontestat, iar, după cum afirmă și autoarele în prezentarea volumului, lucrarea constituie o premieră pentru domeniul RLS, ceea ce face să nu fie "un produs definitiv, ci o propunere, din nenumăratele posibile, care se cere continuată și perfecționată"<sup>202</sup>.

#### 2.2.7. Descrierea minimală a limbii române ca limbă străină. A1, A2, B1, B2

De dată recentă este volumul *Descrierea minimală a limbii române ca limbă străină. A1, A2, B1, B2*<sup>203</sup>, realizat la Departamentul de limbă, cultură și civilizație românească de la Facultatea de Litere a Universității Babeș-Bolyai din Cluj-Napoca. Lucrarea se evidențiază prin faptul că oferă o descriere, pe toate palierele, a limbii vorbitorului la nivelurile A1, A2, B1 și B2. Aceasta a apărut ca răspuns la nevoile profesorilor care predau RLS/RLNM:

<sup>201</sup> Pentru detalierea acestor teme, vezi Victoria Moldovan, Liana Pop, Lucia Uricaru, *op. cit.*, 2001, p. 104.

<sup>&</sup>lt;sup>199</sup> Victoria Moldovan, Liana Pop, Lucia Uricaru, op.cit., 2001, p. 12.

<sup>&</sup>lt;sup>200</sup> *Ibidem*, p. 13.

<sup>&</sup>lt;sup>202</sup> Victoria Moldovan, Liana Pop, Lucia Uricaru, op. cit., 2001, p. 14.

<sup>&</sup>lt;sup>203</sup> Elena Platon, Ioana Sonea, Lavinia Vasiu, Dina Vîlcu, op. cit., 2014.

"Majoritatea colegilor cu care am interacționat în acest context au semnalat sau au recunoscut lipsa de consens în ceea ce privește profilul fiecărui nivel, afirmând că nici măcar din punctul de vedere al claselor lexico-gramaticale nu le este întotdeauna clar cum ar trebui făcută repartiția pe niveluri, ca să nu mai vorbim de lipsa totală de control asupra lexicului sau a actelor de vorbire aferente fiecărui palier."  $^{204}$ 

Fiecare nivel conține 3 secțiuni: funcțiile comunicative, construcția comunicării (unde sunt incluse clasele gramaticale, elementele lexicale, exprimarea circumstanțelor, construcția enunțului și realizarea coeziunii și a coerenței în textul vorbit) și tipurile de texte. Așadar, în Descrierea minimală își găsesc locul și sferele lexicale, repartizate pe niveluri de competență lingvistică, cu câteva elemente reprezentative pentru fiecare temă, ilustrativ pentru ordinea în care se introduce vocabularul la fiecare nivel. Ca în cazul celorlalte materiale prezentate în secțiunea aceasta, în selecția cuvintelor și în ierarhizarea lor, și în cazul acestui material, autoarele s-au ghidat după:

"manualele de RLS-RLNM existente pe piață, vocabularul minimal al limbii române, dar au contat, cel puțin la fel de mult, experiența de predare a RLS-RLNM și, bineînțeles, intuiția. Aceasta, la rândul ei, s-a ghidat mai ales după criterii precum *frecvența*, *utilitatea*, *productivitatea* sau *puterea derivațională* a unui cuvânt, în funcție de care acesta a fost plasat mai sus sau mai jos pe scala celor patru niveluri"<sup>205</sup>.

Deși constituie cel mai recent și mai complet material, din punctul nostru de vedere, nu ne rămâne totuși decât să subliniem încă o dată nevoia de validare a acestor inventare/selecții pe baza producțiilor studenților nonnativi de limba română.

## 2.2.8. Încercare de plasare pe nivel

Pe lângă acest material valoros, la departamentul menționat, s-a încercat, în anul 2011, realizarea unui inventar lexical al RLS, pornind de la cuvintele din DEX. Având deja o primă plasare a cuvintelor de nivel C2, axată pe criteriul frecvenței, realizată cu ajutorul unui program informatic, o echipă de profesori cu experiență în predarea românei ca limbă străină au verificat repartiția cuvintelor și au încercat să identifice nivelul pentru celelalte cuvinte din lista inițială. Din păcate, în urma

<sup>&</sup>lt;sup>204</sup> Elena Platon, op. cit., 2014, p.4.

<sup>&</sup>lt;sup>205</sup> *Ibidem*, p. 8.

plasării automate, rezultatele au fost surprinzătoare: cuvinte precum *fabrică*, *file* sau *material* au avut un număr de ocurențe (reprezentări) mult mai mare decât *floare* sau *calculator*, cuvinte considerate cu frecvență ridicată în română. Desigur că au existat și situații în care așteptările au fost confirmate, de pildă, în cazul unor cuvinte ca *mamă* sau *a fi*. În urma acestor rezultate contradictorii, s-a ajuns la concluzia că sunt necesare și alte criterii care să confirme apartenența unui cuvânt la un anumit nivel de competență lingvistică.

#### 2.2.9. Fondul principal lexical al limbii române

Ca material reprezentativ pentru starea actuală a domeniului, am decis să reținem și să prezentăm, în cadrul lucrării noastre, și *Fondul principal lexical al limbii române*<sup>206</sup>, de Graur, deoarece reprezintă "partea esențială a vocabularului unei limbi, caracterizată printr-o mare stabilitate, cuprinzând toate cuvintele cu mare frecvență care denumesc, de obicei, noțiuni fundamentale și care sunt, în general, cuvinte vechi, cu numeroase derivate și cu multe expresii și locuțiuni"<sup>207</sup>. El se diferențiază de masa vocabularului tocmai datorită derivatelor pe care le dă în cadrul celei din urmă categorii a vocabularului unei limbi. Presupunem că o mare parte dintre cuvintele din fondul principal lexical se vor regăsi și în inventarul nostru, însă acest lucru urmează a fi confirmat în cea de-a doua parte a lucrării noastre. Mai mult, decizia noastră a fost susținută și de afirmația autorului privind alcătuirea fondului principal al limbii române, studierea manualelor de limbă străină pentru începători, deoarece aici nu intră decât cuvintele des utilizate în limbă<sup>208</sup>.

Lucrarea conține un număr de 1419 cuvinte, considerate de autor ca făcând parte din fondul principal lexical al limbii române. Odată cu această lucrare s-a realizat pentru prima dată, în cazul limbii române, distincția între fondul principal lexical și celelalte "compartimente" ale vocabularului<sup>209</sup>. În selecția sa, lingvistul s-a folosit de criteriile lui Vahmistrov: 1. "Să fie folosit de întregul popor cu sensul lor fundamental concret (sensurile secundare nu fac parte din fondul principal), 2. Să fie vechi, deși există și cuvinte noi care pătrund în fondul principal; 3. Să poată fi folosite pentru a exprima fiecare mai multe noțiuni concrete; 4. Să servească drept bază pentru formarea de derivate., 5. Să intre ca element de bază în expresii uzuale."<sup>210</sup>. Deși preia criteriile, autorul face o serie de observații cu privire la

<sup>&</sup>lt;sup>206</sup> Al. Graur, Fondul principal al limbii române, București, Editura Științifică, 1957.

<sup>&</sup>lt;sup>207</sup> Dictionarul explicativ al limbii române, Bucuresti, Univers Enciclopedic, 1998, p.309.

<sup>&</sup>lt;sup>208</sup> Al. Graur, op. cit., 1957, p. 23-24.

<sup>&</sup>lt;sup>209</sup> *Ibidem*, p.3.

<sup>&</sup>lt;sup>210</sup> Al. Graur, op. cit., 1957, p. 9.

acestea. Pe de-o parte, subliniază faptul că limba trebuie privită în dinamica sa, iar cuvintele, în cadrul limbii literare, și nu al graiurilor locale, iar, pentru ca un cuvânt să poată fi inclus în fondul principal lexical, acesta trebuie să fie cunoscut și folosit intens de toate categoriile populației.

Putem afirma, fără nicio reținere, faptul că, odată cu evoluția societății, are loc și o schimbare în cadrul fondului principal lexical al unei limbi. Acest lucru se datorează modificării centrului de interese al respectivei societăți sau faptului că "ceea ce era de importanță vitală pentru societate în evul mediu nu mai este la fel în secolul al XIX-lea"<sup>211</sup> sau în momentul redactării acestui studiu. România a schimbat un regim politic cu un altul, iar această schimbare a adus și unele modificări în cadrul categoriilor sociale, în domeniul activităților acestor clase ale populației, iar, odată cu fenomenul globalizării, limba a suferit din nou modificări deloc neglijabile.

## 2.2.10. Vocabularul reprezentativ al limbilor romanice

Vocabularul reprezentativ al limbilor romanice este definit drept "lexicul selectat care cuprinde cuvintele cele mai importante din fiecare limbă romanică. [...] un nucleu, care se detasează de masa vocabularului"<sup>212</sup>.

Pentru a selecta elementele lexicale care îl compun, autorii lucrării au aplicat criteriul bogăției semantice cuvintelor din Dicționarul explicativ al limbii române. Dintre acestea, au fost reținute doar acelea care au minimum cinci sensuri proprii sau figurate. Cu ajutorul criteriului puterii de derivare, au fost selectate cuvintele care au cel puțin trei cuvinte derivate pe teren românesc (prin prefixare, sufixare sau derivare regresivă). Spre deosebire de criteriile care au stat la baza selecției făcute de către Graur în cartea sa, lucrarea în discuție se remarcă prin faptul că utilizează în stabilirea vocabularului reprezentativ niște principii mai "largi", mai permisive. Mai mult, aceasta adaugă un nou criteriu, cel al uzajului. Pentru îndeplinirea acestei ultime "condiții", autorii au apelat la dicționarul de frecvență al lui Juilland. De aici, au fost extrase primele 2000 de cuvinte, iar, dintre acestea, au fost excluse 28 de cuvinte care erau învechite sau chiar iesite din uz.

Aplicând cele trei criterii, s-a stabilit vocabularul reprezentativ al limbii române contemporane (VRR), care conține un număr de 2581 de cuvinte, spre deosebire de cele 1419 conținute de fondul principal lexical.

1 4 1

<sup>&</sup>lt;sup>211</sup> Al Graur, op. cit, 1957, p. 21.

<sup>&</sup>lt;sup>212</sup> Marius Sala (coord.), *Vocabularul reprezentativ al limbilor romanice*, Editura Științifică și Enciclopedică, București, 1988, p. 11.

Chiar dacă lucrarea ca atare nu constituie un instrument pentru însușirea și evaluarea lexicului RLS, aceasta poate reprezenta o bună sursă pentru cuvintele internaționale sau cel puțin pentru cele mai cunoscute dintre ele, care sunt și similare ca formă și sens printre limbile romanice, în cazul în care grupul-țintă ar fi restrâns la vorbitori ai acestui grup de limbi.

#### 2.2.11. Romanian Frequency List

Szabo este un alt lingvist care, recent, în cursul anului 2015, a propus o listă de frecvență pentru limba română, pe baza corpusului ROMBAC<sup>213</sup>, în scopul utilizării acesteia în elaborarea testelor de vocabular pe niveluri pentru limba română, pe care acelasi autor le propune (după modelul Vocabulary Levels Tests, în acceptiunea lui Nation). Acesta sustine utilizarea unei liste de 10.000 de elemente, cele mai frecvente cuvinte din corpusul ROMBAC, drept instrument utilizat în predare și evaluare, pe de-o parte, și ca sursă pentru propunerea testelor de vocabular pentru limba română. În cadrul studiului său, *Introducing a Romanian Frequency* List and The Romanian Vocabulary Levels Test. Current Issues<sup>214</sup>, Szabo prezintă metodologia care a stat la baza elaborării acestei liste de frecventă si oferă câteva informații în legătură cu rezultatele obținute. Se arată că aproximativ 200 din primele 5000 de cuvinte din Romanian Word List (RWL) au indicele de frecvență extrem de ridicat, fiind vizibilă o diferentă majoră între distributia acestor cuvinte spre deosebire de cea a celorlalte elemente lexicale. Conform autorului, după primele 500 de cuvinte, indicele de frecventă se echilibrează, după care scade gradual, tendintă predictibilă, putem spune, pe baza surselor bibliografice care tratează această temă. Așadar, ce e important să reținem din concluziile lui Szabo este faptul că primele aproximativ 500 de cuvinte, din punctul de vedere al frecventei, asigură o acoperire destul de mare a textelor. Printre primele cuvinte înregistrate de acesta se numără: de, și, pentru, pe, eu, vrea<sup>215</sup>. E interesant că la rangul 5000 se află cuvinte precum: card, jurnalist și mediocru<sup>216</sup>, cuvinte care, pe baza intuiției și a experienței, am avea

<sup>&</sup>lt;sup>213</sup> Realizat la Institutul pentru Inteligență Artificială, Iași. Acesta este cel mai mare corpus adnotat pentru română și conține cinci părți egale de texte, reprezentând genurile: a. jurnalistic (știri și editoriale), b. juridic, c. ficțiune (romane și poezie scrise sau traduse), d. medicină și farmacie (texte scurte), e. biografii ale diferitelor personalități românești și lucrări de critică literară.

<sup>&</sup>lt;sup>214</sup> Cz. Szabo, "Introducing a Romanian Frequency List and the Romanian Vocabulary Levels Test", în *Current Issues in Linguistic Variation: The 14th international conference of the Department of Linguistics*, Vol. 2, University of Bucharest: Bucharest University Press, 2015, pp. 303–317.

<sup>&</sup>lt;sup>215</sup> Cz. Szabo, op. cit., 2015, p. 310-311.

<sup>&</sup>lt;sup>216</sup> *Ibidem*, p. 311.

tendința să le introducem în cadrul cursurilor de RLS mult mai devreme din mai multe motive.

Pe de-o parte, vorbim despre cuvinte internationale, ceea ce presupune că pot fi însușite mai repede. Ne referim aici la cuvinte internaționale (cognates) care, conform lui Webb și Națion, reprezintă elementele cel mai simplu de însusit în L2. deoarece au aceeasi formă și acelasi conținut ca în L1<sup>217</sup>. Si Szabo, într-un alt studiu<sup>218</sup>, a fost interesat de relatia care se stabileste la nivel cognitiv între lexiconul mental din limba maternă și din celelalte limbi pe care un individ le învață. În acest sens, acesta a desfășurat un studiu de caz cu studenți care au ca L1 limba maghiară, ca L2, româna, și ca L3, engleza. Autorul a încercat, într-o primă fază, să observe cum se raportează studentii la aceste cuvinte internationale. Concluziile sale au fost surprinzătoare. Pe de-o parte, aceste cuvinte au fost recunoscute indiferent că vorbim despre L2 sau L3<sup>219</sup>, apoi, s-a confirmat faptul că nu este nevoie de niciun fel de explicații pentru a identifica și înțelege aceste cuvinte internaționale, în special în cazul adultilor aflati la nivelurile intermediar sau avansat în studiul limbii engleze. Se pare că recunoasterea acestor cuvinte este mai degrabă un proces intrinsec<sup>220</sup>. O altă concluzie relevantă pentru domeniul achiziției limbilor este aceea că, indiferent de similitudinile existente între L2 și L3, rezultatele analizei sale identifică o legătură foarte strânsă între L1 și L3. Astfel, "răspunsurile studenților sugerează că aceștia, când vine vorba despre citire si scriere, conectează cuvintele din lexiconul lor mental de conceptele din limba lor maternă și nu accesează activ cunoștințele din limba a doua"221. Ceea ce ne demonstrează că, desi în cazul cuvintelor internaționale, apelul la L2 are loc, în general, procesul prin care sunt recunoscute cuvintele nu e unul constient, ci, mai degrabă, unul realizat automat, situatie care confirmă asa-numitul model Parasitic al lui Hall<sup>222</sup>. Atât Bot<sup>223</sup>, cât și Biriș sunt preocupate de această temă, considerând că limba română are un caracter particular, în cazul achiziției L2,

<sup>&</sup>lt;sup>217</sup> Stuart Webb, Paul Nation, *How Vocabulary is Learned*, Oxford, Oxford University Press, 2017, p. 38.

<sup>&</sup>lt;sup>218</sup> Csaba Z Szabo, "Exploring the Mental Lexicon of the Multilingual: Vocabulary Size, Cognate Recognition and Lexical Access in the L1, L2 and L3", în *Eurasian Journal of Applied Linguistics* 2 (2) (2016), pp. 1-25.

<sup>&</sup>lt;sup>219</sup> Csaba Z Szabo, op. cit, 2 (2) (2016), p.20.

<sup>&</sup>lt;sup>220</sup> *Ibidem*, p.21.

<sup>&</sup>lt;sup>221</sup> Ibidem.

<sup>&</sup>lt;sup>222</sup> C. J. Hall, "The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development", în *International Review of Applied Linguistics in Language Teaching*, 2002, 40, pp. 69–87, disponibil online: http://doi.org/cntmwz. si Csaba Z Szabo, *op. cit.*, 2 (2) (2016), p.5.

<sup>&</sup>lt;sup>223</sup> Livia Bot "Dificultăți generate de sistemul morfosintactic al limbii române", în *Logos și methodos*, 1982, p. 95 *apud* Gabriela Biris, *op. cit.*, 2012, p. 180.

cunoscând "o dublă raportare: la limba maternă și la cel puțin încă o limbă străină"<sup>224</sup>, pe care studenții o dețin deja, desigur, la niveluri diferite. Însă acest aspect ar trebui considerat un atu și folosit ca atare, prin introducerea cuvintelor internaționale, în special cuvinte de origine engleză, căci aceasta e limba de circulație internațională cunoscută de cei mai mulți studenți, cum e și situația noastră (vezi *infra* cap. 6)<sup>225</sup>.

Pe de altă parte, e important să ținem cont de grupul nostru țintă. Cu toate că rezultatele lui Szabo sunt fără îndoială utile și relevante pentru limba română ca L1, probabil și pentru româna ca limbă străină, studiată în mediu nevorbitor, în situația noastră, RLS în mediu vorbitor, considerăm că o listă bazată doar pe mostre de limbă română ca limbă maternă pot conduce spre rezultate eronate sau cel puțin nu foarte relevante, care, cu siguranță, nu vor ilustra cu adevărat situația. Mai mult, este esențial să reținem că această listă a fost elaborată doar pe baza unui corpus care nu poate fi considerat reprezentativ pentru limba română, deoarece nu include și limbă vorbită sau elemente lexicale specifice vieții de zi cu zi. Având în vedere că mostrele de limbă selectate sunt extrem de specializate, considerăm că acest corpus și deci și materialele obținute pe baza acestuia ar putea constitui un ajutor pentru achiziția L2 și, implicit, pentru însușirea vocabularului aparținând nivelurilor superioare, B2-C2, niveluri care, pe de altă parte, nu ar avea o atât de mare nevoie de o repartiție pe niveluri.

Cu toate acestea, este interesant de reținut concluzia autorului, formulată după ce a confruntat primele 3000 de cuvinte din lista sa de frecvență cu secțiunea de literatură din corpusul ROMBAC. Acesta a descoperit că, "știind primele 2000-3000 de cuvinte din română, un individ poate demonstra înțelegerea a patru din cinci cuvinte dintr-un text" [tr.n.]<sup>226</sup>. Considerăm că o validare mai relevantă a acestei liste ar fi fost util de realizat și prin confruntarea RWL cu alte texte literare, care nu au fost incluse în corpusul ROMBAC, altfel, considerăm că aceste concluzii sunt valabile doar pentru aceste mostre de limba română. Va fi interesant de observat, din punct de vedere științific, diferențele obținute dintre rezultatele lui Szabo și inventarul nostru lexical, bazat pe corpusul vorbitorilor nonnativi, întrucât "o analiză fundamentată numai pe frecvență, în cazul vorbitorilor nativi nu reține anumite cuvinte care le sunt folositoare cursanților L2 și care au o frecvență mare în vocabularul folosit în clasă." [tr. n.]<sup>227</sup>.

<sup>&</sup>lt;sup>224</sup> Livia Bot "Dificultăți generate de sistemul morfosintactic al limbii române", în *Logos și methodos*, 1982, p. 95 *apud* Gabriela Biriș, *op. cit.*, 2012, p. 180.

<sup>&</sup>lt;sup>225</sup> Gabriela Biris, op. cit., 2012, p. 180.

<sup>&</sup>lt;sup>226</sup> Cz. Szabo, op. cit., 2015, p. 311.

<sup>&</sup>lt;sup>227</sup> Cambridge ESOL: Research notes, Issue 41, August 2010, Cambridge University Press, p. 3.

Se cuvine să amintim aici și o cercetare care se desfășoară la Universitatea București de către Gabriela Biriș și Alexandru Anghelina și care are în centru analiza manualelor de predare a românei ca limbă străină cu scopul de a identifica elementele lexicale specifice fiecărui nivel<sup>228</sup>.

Desigur că există și alte materiale didactice, în special suporturi de curs<sup>229</sup>, care conțin, în cadrul unităților sau la finalul fiecărei secțiuni, scurte liste de cuvinte și expresii utile, cu siguranță, în procesul de predare-învățare-evaluare al RLS, dar care au fost constituite pe aceleași principii, ale intuiței și ale experienței. După prezentarea stadiului actual al cercetării, în capitolul următor ne îndreptăm atenția asupra motivelor pentru care am ales această temă de cercetare, evidențiind principalele avantaje pe care inventarul lexical al RLS, pentru nivelul A1 le poate aduce domeniului limbii române ca L2.

<sup>&</sup>lt;sup>228</sup> O prezentare a rezultatelor obținute până la acest moment a fost făcută la Conferința internațională Discurs polifonic în româna ca limbă străină (RLS), 29 octombrie 2021, Universitatea Babeș-Baolyai, Cluj-Napoca, unde autorii au susținut comunicarea cu titlul Conținut lexical în manuale RLS – o perspectivă actualizată.

<sup>&</sup>lt;sup>229</sup> Reţinem, aici, câteva exemple de lucrări relevante din diferite centre de predare și evaluare a RLS: Liana Pop, Victoria Moldovan, Ileana Mureșanu, *Româna în 15 lecții. Curs intensiv de limba română*, Cluj-Napoca, Universitatea din Cluj-Napoca, 1989; Stela Zdrenghea, *Limba română. Nivel A (începători)*, Cluj-Napoca, Universitatea Babeș-Bolyai, 1995; Ana Dorobăţ, Mircea Fotea, *Româna de bază. Essential Romanian (2 vol.)*, Iași, Institutul European, 1999; Grigore Brâncuș (coord.), Adriana Ionescu, Manuela Saramandru, *Limba română: Manual pentru studenții străini* (ed. a 5-a), București, Editura Universității din București, 2003; Liana Pop, *Româna cu sau fără profesor. Le Roumain ou sans professeur. Româna:* A1-B1, Cluj, Editura Echinox, 2008; Liana Pop (coord.), *Seria Autodidact. Româna:* A1-B1. Caietul Studentului, Cluj, Editura Echinox, 2008; Daniela Kohn, Puls. *Manual de limba română pentru străini. Curs / Exerciții*, Iași, Polirom, 2009; Platon Elena (coord.), *Româna ca limbă străină (RLS). Caiete didactice. A1+*, Cluj-Napoca, Editura Casa Cărții de Știință, 2012; Cristina-Valentina Dafinoiu, Laura-Elena Pascale, *Limba română. Manual pentru studenții străini din anul pregătitor (Nivel A1-A2*), ediția a 3-a, București, Editura Universitară, 2015.

# CAPITOLUL III – DESPRE NECESITATEA ȘI RELEVANȚA INVENTARELOR LEXICALE<sup>230</sup>

În încercarea de a observa și de a descrie limbile străine din cât mai multe perspective, în prezent, atentia cercetătorilor se îndreaptă înspre descrierile complete realizate pentru diferite niveluri de competentă lingvistică, unde un loc important este ocupat de problematica elaborării inventarelor lexicale. Precum în cazul altor documente lingvistice europene de largă circulație, și în acest caz, s-a elaborat un document European, ghidul Reference Level Descriptions for National and Regional Languages (RLD). Guide for the production of RLD<sup>231</sup>, unde sunt evidentiate etapele în vederea descrierii complete a unei limbi. Acest ghid are rolul de a oferi un cadru general în care să se înscrie toate aceste descrieri, în scopul asigurării unității și a continuității conceptelor și a ideilor promovate prin documentele lingvistice europene existente la această oră. Întrucât este cunoscut faptul că lexicul unei limbi se află într-o continuă transfomare, iar granitele dintre niveluri sunt foarte fine si foarte usor de depășit, inventarele lexicale au fost incluse în categoria denumită inventare "necesare și deschise"<sup>232</sup>. Extrem de impresionant este faptul că, deja, conform paginii dedicate RLD, peste 20 de limbi dispun de astfel de descrieri – din perspective diferite, pe unul sau mai multe niveluri - pornind de la RLD, si, din păcate, deocamdată, limba română nu se numără printre acestea, cu toate că utilitatea unui astfel de material nu este negată. Probabil că, în opinia unor specialisti, acest neajuns ar putea fi rezolvat, prin simpla transferare, pe bază de traducere, a unui model deja consacrat, și limbii române.

Cu toate că un asemenea procedeu ar reuși, poate, să acopere o parte dintre lipsurile cercetării din domeniul RLS, considerăm că acest transfer nu este realizabil decât, eventual, la nivelul structurării materialului, al unui tipar ce va trebui umplut cu un alt conținut, conform particularităților limbii române. Făcând referire la pericolul transferului prin traducere a conținuturilor dintr-o limbă într-alta, Platon dă

<sup>&</sup>lt;sup>230</sup> În cadrul acestui capitol am preluat fragmente deja publicate în studiul nostru, intitulat "Observații asupra utilității inventarelor lexicale", în vol. Înspre și Dinspre Cluj: Contribuții lingvistice. Omagiu profesorului G. G. Neamțu la 70 de ani, coord. Ionuț Pomian, ed. Nicolae Mocanu, Cluj-Napoca, Editura Scriptor și Editura Argonaut, 2015, pp. 54-61.

<sup>&</sup>lt;sup>231</sup> Reference Reference Level Descriptions for National and Regional Languages (RLD). Guide for the production of RLD, Version 2, Language Policy Division, Council of Europe, Strasbourg, November 2005.

 $<sup>^{232}</sup>$ Ibidem.

exemplul descrierii din pasaportul lingvistic european, de la nivelul A1, întelegere după citire, unde apare descrierea: "pot să înteleg nume cunoscute, cuvinte și propozitii simple, de exemplu, din anunturi, afise sau cataloage"<sup>233</sup>. Descrierea se dovedește ineficientă în cazul unei expresii idiomatice de tipul "Bună de-i sar capacele", unde, după formă și context, expresia pare că ar apartine nivelului A1, însă, după cum arată autoarea, sensul acesteia nu ar putea fi înteles, întrucât "pentru decodarea corectă a unui asemenea text este nevoie de cel putin un nivel B1 de cunoaștere a limbii, acolo unde se pot realiza deja inferențe și interpretări mai complexe"<sup>234</sup>. Aşadar, se evidentiază, prin acest exemplu, încă o dată, ideea conform căreia "CECRL este un cadru orientativ, general pentru toate limbile, dar stabilirea cu exactitate a diferentelor dintre enunturile mai simple sau mai complicate [în cazul acesta] urmează să fie făcută de practicienii din domeniul fiecărei limbi"<sup>235</sup>.

În scopul racordării la principiile politicii lingvistice europene și al exploatării experienței acumulate în cazul altor limbi, considerăm că înscrierea pe linia acestora este cu atât mai necesară si mai importantă pentru cercetările viitoare desfăsurate în domeniul RLS. Acesta constituie motivul deciziei noastre de a ne apleca atentia asupra lexicului, care, până la această oră, nu beneficiază de o repartiție a vocabularului pe nivelurile comune de referință din CECRL, cu toate că există mulți specialiști care, de-a lungul timpului, au subliniat această necesitate. Desi ne-am propus, initial, să elaborăm inventarele lexicale pe toate nivelurile comune de referință din CECRL, după o analiză a materialelor existente și a corpusului de care dispunem, am decis să ne limităm, pentru moment, cercetarea la nivelul A1, nivel care, din punctul nostru de vedere, are nevoie de o restrângere a lexicului vehiculat în manuale și în celelalte materiale de predare, învățare și evaluare disponibile pe piață, pentru un cât mai bun control al acestuia. În ultimele studii de specialitate, tot mai frecvent apar statistici privind numărul de cuvinte – pe nivel si chiar pe test – pe care un vorbitor ar trebui să le cunoască pentru a putea promova examenul de competență lingvistică pentru un anumit nivel<sup>236</sup>.

Este de la sine înțeles că o astfel de clasificare ar fi extrem de utilă și în cazul limbii române, întrucât ar limita numărul de cuvinte folosite în manuale, permitând o învătare progresivă, mai bine controlată, a lexicului, dar, mai ales, în testele pe

<sup>&</sup>lt;sup>233</sup> Paşaportul lingvistic european, disponibil aici: https://europass.cedefop.europa.eu/ro

<sup>&</sup>lt;sup>234</sup> Elena Platon, op. cit., 2008, p. 310.

<sup>&</sup>lt;sup>236</sup> James Milton, "The development of vocabulary breath across the CEFR levels, A common basis for the elaboration of language syllabuses, curriculum guidlines, examinations, and textbooks across Europe", în Eurosla Monographs series 1 Communicative proficiency and linguistic development, 2010, pp. 211-232.

nivel, asigurând candidaților din diferite centre universitare sau celor care au învățat româna, în varii contexte, șanse egale în momentul susținerii unui test care se bazează pe un inventar lexical pe niveluri.

### 3.1. Despre necesitatea ierarhizării lexicului RLS

În demersul nostru de a sublinia necesitatea și deosebita utilitate a acestui inventar lexical pe niveluri, considerăm că se impune, în primul rând, să enumerăm specialistii români care s-au preocupat de acest subject si care consideră bine-venit un asemenea instrument, pentru o cât mai bună dezvoltare a domeniului pe plan stiințific și didactic, deopotrivă. Astfel, Platon, abordând problematica descrierilor pe niveluri, arată că "distribuirea cât mai unitară a conținuturilor lingvistice pe nivel este indispensabilă pentru a se ajunge la stabilirea acelei mult râvnite ordini «naturale» după care ar trebui să se desfăsoare achizitia oricărei limbi"<sup>237</sup>. Si Biris subliniază, la rândul său, nevoia unui inventar lexical, din perspectiva predării, evidențiind cele două aspecte esențiale în ceea ce privește predarea vocabularului L2: "numărul de cuvinte pe care ar trebui să le conțină fiecare nivel de studiu și ce cuvinte exact ar trebui predate, în funcție de etapele învătării"<sup>238</sup>. Un rol la fel de important este identificat de aceeași autoare și în procesul de evaluare, dar și în ceea ce privește elaborarea materialelor destinate RLS: "un inventar lexical este necesar și din perspectiva evaluării cunoștințelor lexicale ale unui vorbitor, dar și a elaborării materialelor lexicografice cu rol didactic"<sup>239</sup>. Cu altă ocazie, Platon aduce în discuție aceeasi problemă: "o descriere completă ar necesita și o inventariere a vocabularului specific fiecărui nivel, cu o marjă de eroare de aproximativ 25%, întrucât se spune că un text poate fi înteles dacă 75% din vocabular este cunoscut<sup>2240</sup>. În acelasi sens, și Burlacu este de părere că "un inventar lexical pe nivel este important în achiziția unei limbi, asigurând o uniformizare atât la nivelul unei singure instituții, cât și la nivel mai larg (de exemplu, toate instituțiile care oferă servicii lingvistice), pe plan național și european"241. Într-adevăr, considerăm că inventarul lexical al RLS ar

\_

<sup>&</sup>lt;sup>237</sup> Elena Platon, "Reflexii asupra conceptului de interlimbă", în vol. *Înspre și Dinspre Cluj: Contribuții lingvistice. Omagiu profesorului G. G. Neamțu la 70 de ani*, coord. Ionuț Pomian, ed. Nicolae Mocanu, Cluj-Napoca, Editura Scriptor și Editura Argonaut, 2015, p. 536.

<sup>&</sup>lt;sup>238</sup> Gabriela Biris, op. cit., 2012, p. 173.

<sup>&</sup>lt;sup>239</sup> Ibidem.

<sup>&</sup>lt;sup>240</sup> Elena Platon, "Repere pentru delimitarea nivelurilor de competență lingvistică în RLS", în G.G. Neamţu, Ştefan Gencărău, Adrian Chircu, *Limba Română – Abordări tradiționale şi moderne*, Cluj-Napoca, Presa Universitară Clujeană, 2009, p. 504.

<sup>&</sup>lt;sup>241</sup> Diana Burlacu, "Repartizarea lexicului RLNM pe nivelurile CECRL – cu aplicație la ciclul primar", în *Actele Conferinței de lansare a proiectului "Perfecționarea cadrelor didactice din învățământul* 

reprezenta un avantaj major în ceea ce privește gradul de unitate în evaluare - un evaluator, de exemplu, dintr-un centru de limbi dintr-o altă țară va avea aceleași cerințe, pe partea de complexitate a vocabularului, din partea unui candidat, ca și un examinator din România. Acest beneficiu important e subliniat și de către Platon:

"Căci, oricât de dificil sau de riscant ar fi să cataloghezi un cuvânt ca aparținând unui anumit nivel de competență, inventarele lexicale, asemenea celorlalte secțiuni din descrierile lingvistice pe nivel, sunt singurele în măsură să ofere un cadru de referință care să ne scutească de greșeala de a evalua competența lexicală a unui cursant de nivel A2, de pildă, prin raportare la întreg inventarul LŢ, și nu doar la cel care a fost predat și care, în consecință, se cere a fi cunoscut. Numai un asemenea instrument este în măsură să asigure o viziune unitară asupra volumului și a profilului bagajului lexical specific fiecărui stadiu, care, altfel, s-ar preda/învăța/evalua absolut haotic"<sup>242</sup>.

De-a lungul timpului, s-a dovedit că vocabularul poate fi considerat unul dintre cei mai importanți factori, dacă nu chiar cel mai important, în înțelegerea unui text citit/ascultat/scris sau vorbit<sup>243</sup>. De pildă, Laufer și Ravenhorst-Kalovski au descoperit că 64% dintre diferențele apărute în cazul rezultatelor la testele de citire pot fi puse pe seama vocabularului<sup>244</sup>. La o concluzie asemănătoare a ajuns și Milton, în studiul său dedicat extinderii vocabularului pe nivelurile CECRL<sup>245</sup>. Acesta susține că "60, până la 70%, dintre diferențele apărute la examinările pe nivelurile CECRL pot fi explicate prin diferențele în dimensiunea vocabularului de la un nivel la altul" [tr.n.]<sup>246</sup>. Această observație a rezultat în urma analizei datelor obținute cu privire la engleza ca limbă străină, predată în Ungaria și Grecia, franceza ca limbă străină, din Spania și Grecia, și greaca predată în Grecia.

Și Todea<sup>247</sup> aduce în discuție importanța vocabularului pe nivel, în special a celui receptiv, implicat în procesul de înțelegere a textelor scrise. Deși Alderson a

preuniversitar care predau limba română minorităților naționale", Cluj-Napoca, Editura Casa Cărții de Știință, 2011, p. 240.

<sup>&</sup>lt;sup>242</sup> Elena Platon, op. cit., 2015, p. 536.

<sup>&</sup>lt;sup>243</sup> Pentru o prezentare a specialistilor preocupați de importanța lexicului, vezi Batia Laufer, Geke C. Ravenhorst-Kalovski, "Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension", în *Reading in a Foreign Language*, April 2010, Volume 22, No. 1, pp.15-30.

<sup>&</sup>lt;sup>244</sup> *Ibidem*, p. 9.

<sup>&</sup>lt;sup>245</sup> James Milton, op. cit., 2010, pp. 211-232.

<sup>&</sup>lt;sup>246</sup> *Ibidem*, p 225.

<sup>&</sup>lt;sup>247</sup> Adriana Todea, "Developing an Analysis Grid for B1/B2 Reading Tasks and Items", în *Studia*. *Philologia*, LXII, 2/2017, vol. 62, pp. 167-182.

arătat că înțelegerea unui text poate avea loc dacă 95% dintre cuvinte sunt cunoscute, restul putând fi deslușite din context datorită strategiilor de receptare pe care cititorul și le-a dezvoltat, transferabile în L2<sup>248</sup>, autoarea subliniază că, în această situație, vorbim despre o lectură atentă a textului. Totuși, lucrurile stau diferit atunci când este vorba despre alte tipuri de înțelegere, care vizează anumite părți ale textului, de exemplu, așa-numita "frunzărire" a textului (*skimming*), sau lectura în diagonală (*scanning*). În cazurile acestea, susține Todea, avem nevoie ca atât vocabularul, cât și structurile gramaticale utilizate să fie la nivelul urmărit, neluând în considerare procentajul dat de Anderson, deoarece ne referim strict la anumite părți din text care trebuie să fie înțelese (titlu, subtitlu, introducere, concluzie etc.<sup>249</sup>), iar, dacă cititorul se întâlnește aici cu cele 5% dintre elementele de vocabular sau dintre structurile gramaticale peste nivel, înțelegerea textului nu va avea loc<sup>250</sup>.

Considerăm extrem de pertinentă și de utilă constatarea de mai sus, întrucât relevă importanta unui inventar lexical pe niveluri care ar putea constitui un real ajutor si în etapa de selectie a textelor propuse pentru citire. La fel cum se întâmplă si în cazul Text Inspector sau Compleat Word Tutor (prezentate în capitolul 2), si în situația disponibilității inventarului pe niveluri s-ar putea adapta textele-input, în funcție de nivelul de înțelegere al studenților. Si Nation, în studiile sale dedicate receptării textului scris și, implicit, vocabularului receptiv, arată că "dacă un curs sau un text contine prea multe cuvinte necunoscute pentru cel care învată, atunci acesta va face un efort mult prea mare pentru a-si explica aceste cuvinte încât va pierde esenta textului" [tr. n.]<sup>251</sup>. Asadar, conform acestei opinii, dacă un text conține unul sau două cuvinte necunoscute pe fiecare rând, atunci acesta devine prea dificil de înțeles pentru oricine. Atât Hu și Nation, cât și Carver susțin că e nevoie de o acoperire de 98-99% a vocabularului pentru a se putea asigura înțelegerea textului, dar și pentru a-i asigura cititorului o lectură plăcută<sup>252</sup>. Dacă textul conține prea multe cuvinte necunoscute, se va pierde scopul activității, transformând-o, dintr-una dedicată dezvoltării fluentei în lectură, într-una de învătare a vocabularului<sup>253</sup>.

. .

<sup>&</sup>lt;sup>248</sup> J. Charles Alderson, *Assessing Reading*, Cambridge University Press, 2000, p. 35.

<sup>&</sup>lt;sup>249</sup> În accepțiunea lui J. Cyril Weir, Yang Huizhong, Jin Yan, Studies in Language Testing 12: An empirical investigation of the componentiality of L2 Reading in english for academic purposes, Cambridge University Press, 2000, p. 40 apud Adriana Todea, op. cit., p. 170.

<sup>&</sup>lt;sup>250</sup> După Adriana Todea, op. cit., p. 170.

<sup>&</sup>lt;sup>251</sup> După I.S.P. Nation, op. cit., 2016, p. 5.

<sup>&</sup>lt;sup>252</sup> Rob Waring, Paul Nation, "Second language reading and incidental vocabulary learning", în *Angles on the English- speaking world*, vol. 4, 2004, p. 108.

<sup>&</sup>lt;sup>253</sup> I.S.P. Nation, J. Deweerdt, "A defence of simplification", in *Prospect* 16(3), 2001, pp. 55-67.

Pe lângă importanța cunoașterii nivelului, Todea susține că "foarte importantă este și sfera tematică din care face parte tema textului. În cazul în care avem de-a face cu o temă necunoscută anterior, vocabularul vehiculat în text, cu siguranță, îi va fi necunoscut candidatului" [tr. n.] <sup>254</sup>. Or, o asemenea problemă ar putea fi, din nou, preîntâmpinată, prin stabilirea clară a sferelor tematice adecvate fiecărui nivel de referintă, odată cu realizarea unui inventar pe nivel. Ducând mai departe analiza sa, autoarea observă că "propunerea unei scale potrivite care să releve cu adevărat nivelul cerut în cazul cunoștințelor gramaticale și de vocabular cu rol în întelegere nu e unul dintre punctele forte ale CECRL" [tr. n.]<sup>255</sup>. Suntem întru totul de acord cu această afirmație, amintind că, în CECRL, competența lingvistică a fost usor neglijată – dispune de putini descriptori, iar cei existenti au un caracter general, fiind axați mai ales pe producere -, deși e cunoscut faptul că există diferențe considerabile între competentele lingvistică, de receptare și de producere. După o analiză critică a grilei generale pentru citire și a celei pentru volumul vocabularului, Todea a evidentiat o serie de inconsecvente în mentionarea aspectelor de ordin lingvistic, dar si incongruenta informatiilor din cele două grile<sup>256</sup>. Încercând să uniformizeze informatiile, autoarea propune o scală destinată vocabularului necesar în citire, făcând diferența între niveluri prin utilizarea unor nuanțatori de sens, astfel:

Tabelul 5 – Scala propusă de către Todea pentru vocabularul necesar în citire

A2	Poate înțelege vocabularul cu grad ridicat de frecvență din texte, constând în
	informații simple, familiare și concrete.
B1	Poate înțelege vocabularul frecvent din texte constând în teme complexe, dar
	familiare, fapte.
B2	Poate înțelege o gamă largă de cuvinte, cu frecvență redusă, dar și expresii idiomatice
	cu frecvență ridicată din texte, constând în fapte complexe, idei, opinii pe teme
	familiare. <sup>257</sup>

Așadar, prin îndreptările propuse, autoarea consideră că o scală de succes, care ar viza vocabularul din receptarea scrisă, ar trebui construită urmărind extinderea sferelor tematice, pe de-o parte, și includerea vocabularului în funcție de frecvența sa în comunicare, pe de altă parte. Mai mult, Todea susține că un set de descriptori pentru vocabular ar trebui să fie generați în funcție de combinarea și de corelarea a patru aspecte esențiale: "folosirea în comunicare (frecvența), domeniile lingvistice (general vs idiomatic, general vs specializat), cunoștințele anterioare

<sup>&</sup>lt;sup>254</sup>Adriana Todea, op. cit, p. 172.

<sup>&</sup>lt;sup>255</sup> Ibidem.

<sup>&</sup>lt;sup>256</sup> *Ibidem*, p. 173.

<sup>&</sup>lt;sup>257</sup> Grilă tradusă din Adriana Todea, *op. cit.*, p. 173.

despre temă (familiaritatea cu tema) și complexitatea conținutului (abstract vs factual, complexitatea informațiilor)" [tr. n.]<sup>258</sup>. Considerăm această corelație extrem de importantă, fiind absolut natural ca locutorii să fie puși în fața unor fapte de limbă care se încadrează coerent în evoluția lor lingvistică.

Pe plan internațional, Nation argumentează necesitatea listelor de cuvinte, mai ales ca instrumente de bază în dezvoltarea unor materiale eficiente în domeniul achizitiei limbilor. Listele de cuvinte, sustine acesta, stau la baza structurării unui bun curs orientat asupra vocabularului unei limbi, constituie fundamentul unor materiale valoroase destinate citirii la anumite niveluri (este vorba despre așanumitele graded readers, opere literare sau nu, adaptate pe niveluri), reprezintă punctul de plecare în cercetarea asupra volumului vocabularului, precum si în dezvoltarea testelor de vocabular<sup>259</sup>. Totuși, "pentru ca o listă de cuvinte să fie considerată utilă, aceasta trebuie să fie bine realizată" [tr. n.]<sup>260</sup>, iar, ca această conditie să fie îndeplinită, esentiale sunt unitatea de măsură (the unit of counting) și tipul corpusului aflat la baza listei<sup>261</sup>, aspecte pe care le vom aborda și noi în capitolele următoare. Considerăm extrem de importantă precizarea, de fiecare dată, a scopului pentru care o listă de cuvinte a fost realizată. În plus, Nation sugerează să se specifice si dacă aceasta va fi folosită la producere sau la receptare, dacă este o listă obtinută folosind producții orale sau scrise, precum și nivelul pentru care este destinată<sup>262</sup>. Utilitatea listelor de cuvinte se poate observa în mai multe situatii, precum: proiectarea cursurilor, predarea si învătarea limbilor, elaborarea manualelor sau a altor materiale auxiliare adaptate pe nivel, analiza dificultății vocabularului dintr-un text și, desigur, în evaluare, mai ales în proiectarea testelor de vocabular sau a sectiunilor de vocabular din testele de competentă lingvistică. Nu mai putin importante sunt pentru domeniul lexicografic, domeniu în care Stoica remarcă "lipsa unor instrumente lexicografice construite din perspectiva românei ca limbă străină"263. Autoarea consideră că multe dintre erorile lexico-semantice care pot fi regăsite în discursul celor care învată RLS se datorează consultării unor dictionare bilingve sau de sinonime, unde nici informatia lexicală, nici cea fonetică sau morfologică nu apar clar evidențiate/explicate. În acest context, Stoica subliniază

<sup>&</sup>lt;sup>258</sup> Pentru mai multe detalii, vezi Adriana Todea, op. cit., p. 174.

<sup>&</sup>lt;sup>259</sup> I.S.P. Nation, op. cit., 2016, p. XI.

<sup>&</sup>lt;sup>260</sup> *Ibidem*, p. 3.

<sup>&</sup>lt;sup>261</sup> După I.S.P. Nation, op. cit., 2016, p. 3.

<sup>&</sup>lt;sup>262</sup> I.S.P. Nation, op. cit., 2016, p. 34.

<sup>&</sup>lt;sup>263</sup> Gabriela Stoica, op. cit., 2012, p. 190.

necesitatea realizării unui dicționar de română ca limbă străină<sup>264</sup>, redactat din perspectiva unui utilizator nonnativ, argumentând: "calitatea asimilării unei limbi străine depinde în bună parte și de calitatea, eficiența și exactitatea materialului didactic/lexicografic utilizat"<sup>265</sup>. Unii ar combate această afirmație, arătând că, în cazul RLS, este disponibil dicționarul ASA. Utilitatea sa este incontestabilă, însă acesta nu este destinat exclusiv studenților străini și, mai mult, este bilingv, explicațiile și exemplele bazându-se pe limba engleză. Așadar, acest dicționar presupune ca studentul străin care îl consultă să cunoască limba engleză la un nivel suficient de avansat pentru a înțelege explicațiile și exemplele oferite. Or, conform autoarelor, majoritatea exemplelor aparțin nivelurilor B1-B2 (vezi *supra*, 2.2.4.)<sup>266</sup>.

## 3.2. Perspectiva CECRL asupra lexicului ierarhizat pe niveluri

Și în ceea ce privește selecția lexicului utilizat în predare, învățare și evaluare, găsim, în CECRL, unele recomandări. Autorilor de teste și de materiale de predare li se recomandă să selecteze cuvintele care vor fi incluse în aceste materiale, în scopul asigurării unui grad cât mai ridicat de transparență și de coerență în procesul educațional. Astfel, se sugerează să se selecteze cuvintele și expresiile în funcție de nevoile studentului, respectând principiile lexico-statistice de selecție a elementelor lexicale cu frecvență mai ridicată, dar care să permită și dezvoltarea coerentă a vocabularului, ca răspuns la nevoile studentului, atunci când acesta e angajat în rezolvarea unei sarcini de comunicare<sup>267</sup>.

Din păcate, în ciuda bunelor intenții ale autorilor CECRL, se poate remarca lipsa unor indicații clare privind cuvintele sau temele specifice pentru fiecare nivel. Aceste informații pot fi identificate doar în linii mari, în descriptori, drept recomandări generale destinate utilizatorilor acestui document. Aceștia sunt invitați să examineze și să determine, în funcție de caz, de limba descrisă, "elementele lexicale (locuțiuni idiomatice și cuvinte izolate) pe care elevul va trebui să le

<sup>&</sup>lt;sup>264</sup> Conform autoarei, și Rodica Zafiu a fost preocupată de acest subiect, susținând necesitatea unui astfel de material lexicografic în cadrul colocviului "Limba română în lume. Comunicare, cercetare, export cultural", 28-30 octombrie 2005, București.

<sup>&</sup>lt;sup>265</sup> Gabriela Stoica, op. cit., 2012, p. 190.

<sup>&</sup>lt;sup>266</sup> În context lexicografic, semnalăm două titluri care tratează subiectul creării unui dicționar pentru RLS (DRLS/DERLS). Este vorba despre: Elena Platon, "Un proiect utopic de lexicografie didactică în RLS", în Elena Platon, Cristina Bocoș, Diana Roman, Lavinia Vasiu (editori), *Discurs polifonic în româna ca limbă străină (RLS)*, nr. 2/2020, Editura Presa Universitară Clujeană, 2020, pp. 357-378 și Lorena Sacalîș, "Un model de lucru în DERLS", la Conferința internațională *Discurs polifonic în româna ca limbă străină (RLS)*, 29 octombrie 2021, Universitatea Babeș-Bolyai, Cluj-Napoca (comunicare).

<sup>&</sup>lt;sup>267</sup> CECRL, pp. 150-151.

recunoască sau să le folosească ori cu care va trebui să fie înzestrat pentru a face acest lucru"<sup>268</sup>, pe de-o parte, și să decidă "cum vor fi ele selecționate și clasificate"<sup>269</sup>, pe de altă parte.

Așadar, *CECRL*, prin caracterul său general, aplicabil tuturor limbilor europene, îndeamnă la descrierea vocabularului, ținând cont de patru direcții:

- 1. "selectarea cuvintelor-cheie și a expresiilor, tematic, a) în funcție de nevoile comunicative ale studenților, b) în funcție de specificul cultural, și nu numai, al limbii în cauză:
- 2. selectarea în funcție de frecvența cu care apar în corpusuri sau, din nou, tematic;
- 3. selectarea unor texte autentice, vorbite sau scrise, pentru a se preda/învăța cuvintele pe care le conțin;
- 4. dezvoltarea vocabularului în funcție de nevoile studentului angajat în comunicare, fără a se decide anterior structura"<sup>270</sup>.

Desigur că, din perspectiva predării, oricare dintre opțiunile amintite poate funcționa la fel de eficient, cu toate că, neavând un control asupra elementelor predate (dacă decidem să urmăm direcția numărul patru, de exemplu), va fi greu de asigurat dezvoltarea unitară a vocabularului și a competențelor de comunicare ale studentului sau ale tuturor studenților din aceeași clasă. Dacă avem în vedere însă perspectiva evaluării, pentru a putea asigura unitate și control asupra elementelor lexicale evaluate, este ușor de remarcat că opțiunile 3 și 4 nu satisfac aceste exigențe, întrucât nu presupun nicio metodă de control al elementelor introduse. Considerăm că, mai degrabă, este posibil ca, prin combinarea celorlalte două variante (1 și 2), să ajungem la o situație cât mai apropiată de cea dorită.

Modele valoroase găsim în cazul limbilor de mare circulație, pentru care există deja inventare lexicale. Aici, putem remarca un demers diferit. De pildă, după cum am arătat în cadrul capitolului 2, inventarul limbii engleze, EVP, are mai degrabă un caracter didactic (conține definiții și exemple) și a fost alcătuit din perspectiva a ceea ce este capabil vorbitorul nonnativ să facă la final de nivel, la momentul evaluării, însă ținând cont de frecvența cu care apar aceste cuvinte în corpusul manipulat. Alte inventare au urmat un demers diferit și au fost alcătuite asemenea unor liste simple de cuvinte, fără alte informații suplimentare, realizate cu

<sup>&</sup>lt;sup>268</sup> CECRL, p. 94.

<sup>&</sup>lt;sup>269</sup> Ibidem.

<sup>&</sup>lt;sup>270</sup> Françoise Kusseling, Wilfried Decoo, *Europe language learning: The challenges of comparable assessment*, în European Studies Conference Selected Proceedings, University of Nebraska, Omaha, 2009, p.4.

scopul de "a contribui la dezvoltarea unor programe pentru îmbunătățirea instrumentelor de analiză și descriere a limbii" [tr. n.]<sup>271</sup>.

### 3.3. Inventar lexical vs dictionar

Pentru a ilustra cât mai clar ceea ce reprezintă un inventar lexical, rămânem tot la cazul modelelor oferite de limbile de circulație internațională. În primul rând, este important de reținut că *inventarele lexicale* nu sunt dicționare, ci *liste care arată* progresia introducerii, respectiv, a testării elementelor lexicale într-o limbă străină, în funcție de nivelurile de competență lingvistică. Așadar, scopul lor principal nu este acela de a oferi definiții sau exemple și contexte de utilizare ale cuvintelor decât dacă, încă de la început, se urmărește utilitatea materialului și din perspectiva autodidactului. Din cele afirmate, reținem că inventarele nu sunt dicționare și nici nu se doresc a fi, întrucât ar avea același rol ca cele din urmă, iar uzul lor ar fi redus din cauza volumului și a greutății lor. Fără îndoială că specificarea nivelului de competență lingvistică și în dicționare, pe de altă parte, pornind de la acest inventar lexical, ar fi extrem de utilă și ar veni în sprijinul celor care doresc să își însușească sau să-și perfecționeze cunoștințele de limbă.

Chiar dacă toate dicționarele au la bază un corpus, abia în anul 1980 s-a utilizat tehnologie bazată pe analiza de corpus computerizată pentru a obține un dicționar dedicat L2. Este vorba despre COBUILD English Language Dictionary for learners of English as a foreign language<sup>272</sup>. Acesta a fost și primul dicționar care a folosit simboluri de frecvență pentru a arăta cuvintele cele mai utile. În acest sens, s-a folosit simbolul diamantului, care era cu atât mai plin cu cât era cuvântul mai frecvent. The Longman Dictionary of Contemporary English, pe de altă parte, conținea simbolurile S1, S2, S3, W1, W2, W3, unde S reprezenta vorbire (speech), W, scriere (writing), iar cifrele reprezentau rangul, respectiv, primele 1000/2000/3000 de cuvinte<sup>273</sup>. A se remarca asemănarea cu Dicționarul practic pentru studenți români, unde Păltineanu, intuind importanța majoră a acestor precizări, a folosit un sistem de diferențiere între cuvintele de introdus inițial și restul vocabularului. Așa cum afirmă și Nation, "introducerea acestui tip de informație întrun dicționar este un pas inovativ și util" [tr. n.]<sup>274</sup>, de aceea autorul susține

<sup>271</sup> Plan curricular del Instituto Cervantes http://cvc.cervantes.es.

<sup>&</sup>lt;sup>272</sup> J. M. Sinclair, (Ed), Collins COBUILD English Language Dictionary. London: Collins, 1987.

<sup>&</sup>lt;sup>273</sup> Prezentat în I.S.P. Nation, *op. cit.*, 2016, pp. 176-177.

<sup>&</sup>lt;sup>274</sup> I.S.P. Nation, op. cit., 2016, p. 177.

introducerea simbolurilor pentru a marca frecvența și rangul intrărilor în dictionare<sup>275</sup>.

De remarcat că, în aceste inventare pe niveluri, nu sunt retinute cuvintele propriu-zise, ci doar anumite sensuri, denumite de către Marello "unități lexicale – lexical units –, întrucât un cuvânt, atunci când este inclus la un anumit nivel, nu apare cu toate sensurile odată, ci numai cu acel sens relevant pentru nivelul său" [tr. n.]<sup>276</sup>. În cazul limbii române, un exemplu ar putea fi reprezentat de cuvântul ochi, de importanță imediată și cu o frecvență ridicată în vorbire. De aceea considerăm că ar trebui inclus în inventarul lexical de nivel A1, însă numai cu sensul de organ al vederii (asadar, constituind la nivelul A1 o primă unitate lexicală). O a doua unitate lexicală, cu sensul de fel de mâncare (ouă ochiuri), pe de altă parte, se va testa mult mai târziu, abia la nivelul B1<sup>277</sup>, celelalte sensuri secundare urmând să se introducă și mai târziu, în funcție de frecvența în comunicare și, mai ales, de utilitatea lor pentru vorbitori, abia la nivelurile avansate, C1 si C2, putând fi introduse expresiile idiomatice având ca nucleu acest cuvânt, aducând în scenă alte semnificatii (de exemplu, a face cuiva ochi dulci, a sorbi din ochi pe cineva etc.). Asadar, acest aspect poate reprezenta un alt argument care sustine diferenta dintre inventare si dictionare. În timp ce în dictionare cuvintele apar înregistrate cu toate sensurile, în inventare același cuvânt va putea fi inclus de mai multe ori.

### 3.4. Utilitatea inventarelor lexicale în procesele de predare/învățare/evaluare

Repartiția pe niveluri de competență lingvistică a elementelor lexicale este extrem de importantă pentru că poate juca un rol esențial în asigurarea unui parcurs coerent în procesul de predare/evaluare. Cercetările actuale ne confirmă încă o dată că "vocabularul este cheia atât pentru competențele de receptare, cât și pentru cele de producere" [tr. n.]<sup>278</sup>, iar odată cu dezvoltarea acestuia este de așteptat să se dezvolte si abilitătile de comunicare.

Referindu-ne la rolul pe care îl au listele de cuvinte în domeniul achiziției limbilor constatăm că, cel mai adesea, acestea se află la baza dezvoltării curriculelor. Acest aspect este foarte important, întrucât diferitele cuvinte dintr-o limbă nu apar cu aceeași frecvență în comunicare. Apelăm din nou la limba engleză pentru a ne

<sup>&</sup>lt;sup>275</sup> I.S.P. Nation, op. cit., 2016, p. 177.

<sup>&</sup>lt;sup>276</sup> Carla Marello, "Word lists in Reference Level Descriptions of CEFR (Common European Framework of Reference for Languages", în *Proceedings of the XV Euralex International Congress*, Department of Linguistics and Scandinavian Studies University of Oslo, 2012, p. 330.

<sup>&</sup>lt;sup>277</sup> Conform Platon, Elena, Sonea, Ioana, Vasiu, Lavinia, Vîlcu, Dina, op. cit., 2014.

<sup>&</sup>lt;sup>278</sup> James Milton, *op. cit.*, 2010, p. 212.

susține afirmația. Astfel, în general, în cazul acestei limbi se vorbește despre două cuvinte (*the* și *of*) care apar cu o frecvență extrem de ridicată, chiar în (aproape) fiecare propoziție.<sup>279</sup> Dacă ne referim însă la majoritatea cuvintelor dintr-o limbă "o mare parte dintre acestea pot fi întâlnite doar o dată sau chiar deloc de către un individ pe tot parcursul vietii sale" [tr. n.]<sup>280</sup>.

Conform lui Nation, cuvintele cu frecvență ridicată ar trebui să constituie elementul central al cursurilor de limbi, mai ales la început, bazându-se pe principiul cost-beneficiu, adică studenții ar trebui să obțină cât mai multe beneficii din efortul pe care îl depun învătând, iar, în plus, prin faptul că le sunt prezentate aceste cuvinte extrem de utilizate se asigură că acestea vor fi și mai des întâlnite, ceea ce va implica automat că utilizatorii vor fi expusi la noi sensuri, pe care si le vor însusi<sup>281</sup>. Prin urmare, un inventar al acestor elemente lexicale ar fi un instrument indispensabil în elaborarea de curricule. Dacă ne referim la proiectarea cursurilor, listele de cuvinte au rolul de a stabili obiective de atins în cadrul unui program de studiu. Totusi, e dificil de determinat cu adevărat influenta acestora asupra cursurilor. Chiar și în cazul limbii engleze, de exemplu, Nation sustine că "majoritatea autorilor de cursuri de limbi nu iau în considerare într-un mod sistematic nici cunostințele de vocabular, nici nivelurile" [tr. n.]<sup>282</sup>, în primul rând, fiindcă listele sunt prea lungi, dar și pentru că acest lucru ar necesita prea multă muncă, ar presupune cunoasterea tuturor listelor, a nevoilor studentilor, iar, apoi, alegerea listei potrivite nevoilor studentilor si, desigur, proiectarea cursului încercând excluderea, pe cât posibil, a altor elemente de vocabular, care nu apartin listei, deci care nu sunt la fel de prolifice în limbă. De aceea se preferă tot mai des fragmentarea acestor liste în subliste, în funcție de rangul ocupat în inventarul principal. Un exemplu în acest sens a fost prezentat în capitolul 2, unde era vorba despre EWL. Considerăm că, din perspectiva predării și a învățării, separarea în astfel de liste poate fi utilă, o listă de 100 de cuvinte în locul uneia de 1000 de cuvinte fiind mult mai ușor de gestionat, întrucât nu va fi atât de copleșitoare nici pentru profesor, nici pentru student. Deși nu credem că va fi posibil să fragmentăm lista noastră precum s-a întâmplat în cazul EWL, avem în vedere, unde este posibil, extragerea anumitor categorii de cuvinte și formarea unor liste separate de inventarul lexical.

<sup>&</sup>lt;sup>279</sup> După I.S.P. Nation, op. cit., 2016, p. 3.

<sup>&</sup>lt;sup>280</sup> Ibidem.

<sup>&</sup>lt;sup>281</sup> *Ibidem*, p. 5.

<sup>&</sup>lt;sup>282</sup> *Ibidem*, p. 172.

Cum elementele de vocabular nu se învață doar într-un cadru organizat, ci și accidental (*incidental learning*), cărțile adaptate pe niveluri (*graded readers*<sup>283</sup>) constituie un atu în acest sens, dar și o necesitate în cazul limbii române, care la acest moment, după cunoștințele noastre dispune doar de un singur volum de texte gradate, pentru nivelurile A1-A2<sup>284</sup>. Acestea ar putea constitui și un sprijin pentru student sau pentru orice persoană care învață limba română, întrucât, prin faptul că nu (prea) conțin elemente peste nivel, pot fi ușor citite, iar, în plus, se va ști că toate cuvintele întâlnite sunt importante și utile, de aceea apar în acele cărți. În aceeași idee, Nation susține că o parte importantă a vocabularului unui individ este dobândită prin expunere la text, prin citire, mai ales dacă un număr mare de cuvinte a fost deja însușit. Se arată că, în cazul limbii engleze, de exemplu, dacă sunt deja predate primele 2000 de cuvinte, pentru a le învăța pe cele din cea de-a treia mie, ar fi nevoie de expunerea la un număr de 300.000 de elemente, adică de aproximativ trei romane pe care să le citească studenții pe parcursul anului universitar, ceea ce reprezintă, după calculele autorului, 10 minute de lectură în fiecare zi, cinci zile pe săptămână<sup>285</sup>.

În ceea ce privește utilitatea listelor de vocabular în **evaluare**, s-a dovedit că, la fel ca în cazul predării și al învățării, dacă sunt bine realizate, acestea vor fi extrem de utile în pregătirea testelor de vocabular, spre deosebire de folosirea unor dicționare, metodă care poate avea rezultate înșelătoare<sup>286</sup>. Însă, și în acest caz, condiția este folosirea unui corpus alcătuit corect, relevant și urmărind obiectivul pentru care a fost creat<sup>287</sup>.

Considerăm că un rol cu atât mai important îl ocupă vocabularul în etapa de (auto)evaluare, unde e necesar să existe o unitate în ceea ce privește: "numărul de cuvinte care ar trebui cunoscute pe nivel, iar, apoi, un inventar al acestora"<sup>288</sup>. Kusseling și Decoo sunt doar doi dintre autorii care afirmă nevoia unei unități la nivel instituțional, atât în ceea ce privește predarea, cât și evaluarea. Astfel, candidații pot preîntâmpina problemele care ar putea să apară la un examen, din cauză că au folosit în pregătire un manual sau o programă care nu acoperă toate cerințele pentru acel examen. Tocmai acesta este motivul pentru care considerăm

<sup>&</sup>lt;sup>283</sup> De exemplu, https://erfoundation.org/wordpress/graded-readers/.

Este vorba despre: Elena Platon, Cristina Gogâță, Lavinia-Iunia Vasiu, Anca Ursa, *România în 50 de povești. Limba română ca limbă străină (RLS). Texte gradate A1-A2*, Editura Presa Universitară Clujeană, Cluj-Napoca, 2021.

<sup>&</sup>lt;sup>285</sup> I.S.P. Nation, op. cit., 2016, p. 173.

<sup>&</sup>lt;sup>286</sup> Pentru mai multe detalii, vezi K. C. Diller, *The Language Teaching Controversy*. Rowley, Mass.: Newbury House, 1978 *apud* I.S.P. Nation, *op. cit.*, 2016, p. 8.

<sup>&</sup>lt;sup>287</sup> După I.S.P. Nation, op. cit., 2016, p. 8.

<sup>&</sup>lt;sup>288</sup> Françoise Kusseling, Wilfried Decoo, op. cit., p. 9.

existența unui inventar lexical pe niveluri atât de importantă. Un asemenea instrument, odată disponibil, poate să însoțească specificațiile pentru examene și să stea la baza realizării manualelor și a programelor din centrele de limbi, constituind elementul care asigură unitatea în predare și în evaluare. Kusseling și Decoo afirmă că "necunoașterea chiar și a unui singur cuvânt dintr-un item la un test poate duce la incapacitatea de rezolvare corectă a acestuia, deși candidatul are toate celelelalte competențe necesare dezvoltate" [tr. n.]<sup>289</sup>. Considerăm validă această observație, mai ales în ceea ce privește nivelurile de început (A1 și A2) și în cazul introducerii unor sensuri noi pentru cuvinte deja cunoscute. Și Szabo susține relevanța listelor de frecvență și a testelor în achiziția L2, arătând că "listele de frecvență și testele pot oferi o vastă cantitate de informație privind structura unei limbi, obiectivele de stabilit pentru cei care o învață ca pe L2, instrumentele pentru a atinge aceste obiective, precum și teste cantitative și standardizate pentru profesori ca să poată măsura progresul studentilor" [tr.n.]<sup>290</sup>.

Utilitatea inventarelor lexicale rezultă și din principiul celor patru etape (four strands) al lui Nation, care propune ca, în cadrul unui curs de limbă, să existe patru părți egale: input centrat pe conținut (meaning-focused input), output centrat pe continut (meaning-focused output), învățare centrată pe limbă (language-focused learning) și dezvoltarea fluenței (fluency development)<sup>291</sup>. Învătarea centrată pe limbă reprezintă predarea și învătarea intentionată a vocabularului, situatie în care, din nou, listele de cuvinte ar fi extrem de utile. La fel de folositoare pot fi și în etapa inputului centrat pe conținut, când, în timpul citirii sau al ascultării, studentul poate lua decizia de a învăța sau nu un anumit cuvânt, prin simpla consultare a listei care îi va arăta dacă este un cuvânt cu frecventă ridicată și deci important și relevant pentru el. Nation avansează chiar ideea unui program informatic sau a unui dictionar electronic care, pe baza listelor de cuvinte, l-ar putea ajuta pe cel care învață să facă alegeri bine informate. Astfel, cu ajutorul acestui program, s-ar putea afla dacă un anumit cuvânt se repetă într-un text sau dacă e în general frecvent, ceea ce implică necesitatea de a fi însușit cât mai curând. Rolul etapei dedicate dezvoltării fluenței este acela de a le asigura studenților un cadru în care desfășoară activități de ascultare, vorbire, citire și scriere cu material deja cunoscut, având drept obiectiv principal să crească viteza cu care folosesc limba în activități de producere și de receptare. "Până când studenții nu dețin un vocabular de cel puțin 3000 de cuvinte, preferabil chiar mai multe, materialul utilizat pentru dezvoltarea fluenței e nevoie să

<sup>20</sup> 

<sup>&</sup>lt;sup>289</sup> Françoise Kusseling, Wilfried Decoo, *op. cit.*, p. 9.

<sup>&</sup>lt;sup>290</sup> Cz. Szabo, *op. cit.*, 2015, p. 307.

<sup>&</sup>lt;sup>291</sup> I.S.P. Nation, op. cit., 2016, p. 6.

fie foarte bine controlat din perspectiva vocabularului pentru a se evita întâlnirea de cuvinte noi în cadrul activităților dedicate fluenței" [tr. n.]<sup>292</sup>.

### 3.5. Concluzii

Ca să rezumăm cele expuse, considerăm că un inventar lexical al limbii române ca limbă străină, pe nivelul A1 din CECRL, ar constitui, în primul rând, un valoros instrument lingvistic și didactic. În plus, ar aduce cu sine o multitudine de avantaje atât pentru specialiști, cât și pentru toți cei interesați de studiul RLS, avantaje pe care le prezentăm în continuare.

Aşadar, un inventar lexical al RLS pe nivelul A1 din CECRL:

- ar constitui un element științific important, care ar sta la baza cercetării vocabularului în domeniul achiziției L2, datorită elaborării sale pe baza corpusului de producții scrise și vorbite ale nonnativilor;
- ar asigura uniformizarea procesului de predare/învățare/evaluare, indiferent de cadrul instituțional în care se desfășoară acesta sau de experiența cadrelor didactice implicate. Pornind de la acest material, se vor putea elabora programe analitice, criterii sau grile de evaluare a competențelor orale și scrise;
- ar constitui un suport pentru elaborarea manualelor, a testelor și a altor materiale didactice, organizate pe niveluri de competență lingvistică;
- ar reprezenta un instrument prețios pentru evaluatorii din țară și din străinătate, care se află în ipostaza de a verifica și de a certifica nivelul de competență lingvistică al unui candidat (în special pentru cei care nu au româna ca limbă maternă și care lucrează, de obicei, în afara granițelor);
- ar constitui un instrument extrem de valoros pentru autorii de cărți adaptate pe nivel;
- ar deveni un material accesibil tuturor categoriilor-țintă, în cazul în care ar fi segmentat în subliste, putând constitui și baza elaborării unor cartonașe de cuvinte, de tip *flashcards*, propriu-zise sau electronice, sau chiar a unor aplicații electronice pentru telefon sau calculator pornind de la acestea<sup>293</sup>.

Dacă în cadrul acestui capitol am discutat despre necesitatea și relevanța inventarului lexical al RLS pentru nivelul A1 din CECRL, în capitolul următor ne îndreptăm atenția asupra corpusului și a lingvisticii bazate pe corpus, în general, precum și a necesității unui corpus de producții scrise ale nonnativilor, în special.

<sup>&</sup>lt;sup>292</sup> I.S.P. Nation, op. cit., 2016, p. 177.

<sup>&</sup>lt;sup>293</sup> Detalii despre elaborarea acestor materiale în I. S. P. Nation, *Learning Vocabulary in Another Language.* (Second edition). Cambridge, Cambridge University Press, 2013, capitolul 11.

# CAPITOLUL IV – CORPUSUL LINGVISTIC, ÎN GENERAL<sup>294</sup>

Remarcăm că, tot mai des, se vorbește despre utilitatea corpusului în demonstrarea diferitelor teorii dintr-o varietate de domenii de cercetare. Aria limbilor străine nu face excepție de la această tendință, întâlnind cu o frecvență sporită studii bazate pe corpus (oral sau scris), denumit, în literatura străină de specialitate *learner corpus*, iar, în limba română, *corpus al vorbitorilor nonnativi*. Cercetarea noastră nu ar putea fi dusă la bun sfârșit fără o confirmare din partea lingvisticii bazate pe corpus (*corpus linguistics*).

Corpusul lingvistic reprezintă obiectul de studiu al lingvisticii bazate pe corpus (*Corpus Linguistics*), pe care o vom denumi în continuare LC și care a devenit cunoscută pe scară largă abia în anii 1980<sup>295</sup>. După McEnery și Hardie, domeniul LC este unul eterogen, care se axează pe un set de proceduri și metode pentru studiul limbii. Chiar dacă unele dintre procedurile utilizate sunt extrem de clare, există altele care se află încă în plin stadiu de dezvoltare și care tocmai de aceea pot să ne ajute în rafinarea unor teorii sau în confirmarea altora, care până acum nu se puteau realiza din cauza materialului mult prea extins<sup>296</sup>. Oricum, prin utilizarea celor două mari tipuri de analiză, oferite lingviștilor de LC, analiza calitativă (*concordanțele* – cuvintele văzute în context, și de stânga, și de dreapta) și cea cantitativă (*listele de frecvență*), specialiștii pot găsi răspunsuri la un set de întrebări prestabilite. Așa cum Szudarski<sup>297</sup> afirmă, bazându-se pe argumentele lui Hunston<sup>298</sup>, "se poate susține că LC constituie o abordare pur empirică care se fundamentează pe analize bazate pe frecvență în scopul studierii tendințelor tipice care se regăsesc în limbă privind

<sup>&</sup>lt;sup>294</sup> Fragmente din acest material au fost publicate în studiul nostru, realizat în colaborare cu Vasiu Lavinia-Iunia, intitulat "On the Necessity of a Learner Corpus – Romanian as a Foreign Language (RFL)", în *Globalization and National Identity. Studies on the Strategies of Intercultural Dialogue*, ed. Iulian Boldea, Arhipelag XXI, Târgu Mureş, 2016, pp. 955-966.

<sup>&</sup>lt;sup>295</sup> Michael McCarthy, Anne O'Keeffe, "Historical perspective. What are corpora and how have they evolved", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, p. 5.

<sup>&</sup>lt;sup>296</sup> Tony McEnery, Andrew Hardie, *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge, 2012, p. 1.

<sup>&</sup>lt;sup>297</sup> Paweł Szudarski, *Corpus Linguistics for Vocabulary. A Guide for Research*, Routledge, London and New York. 2018.

<sup>&</sup>lt;sup>298</sup> S. Hunston, *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge, 2002, p. 20, *apud* Paweł Szudarski, *op. cit.*, p. 8.

contextele de ocurență ale anumitor cuvinte"<sup>299</sup>. Mai mult, autorul suține că dacă acceptăm că intuiția vorbitorului nativ nu constituie un factor de încredere în determinarea celor mai frecvente cuvinte dintr-o limbă, atunci trebuie să realizăm că această sarcină va fi și mai complicat de realizat de către un vorbitor nonnativ. Iar în această situație, utilitatea unui corpus al vorbitorilor nonantivi constituie un element-cheie³00. Astfel, tot mai frecvent se recomandă utilizarea corpusului și pentru verificarea și oferirea unor exemple din limba naturală, în defavoarea celor din textele literare.

"Orice teorie lingvistică azi trebuie susținută de o evaluare care să se ancoreze într-un corpus semnificativ de exemple. Corpusurile și metodele de evaluare au devenit atât de semnificative în lingvistica zilelor noastre încât au început să fie organizate importante conferințe internaționale specific orientate pe această problematică"<sup>301</sup>.

Această observație este, desigur, valabilă și în ceea ce privește domeniul RLS. O programă de curs, un manual, un dicționar sau un test, realizate pe baza unui corpus sau validate cu ajutorul analizei acestuia, vor prezenta mai multă încredere decât un material bazat pe intuiția de vorbitor nativ a autorului sau pe experiența de predare a acestuia.

Conform lui Brezina, "spre deosebire de alte surse de informații, precum mitologia, psihologia sau arta, *știința* se bazează pe colectarea sistematică a datelor empirice și pe testarea teoriilor și a ipotezelor" [tr. n.]<sup>302</sup>. Așadar, o afirmație sau o teorie poate fi considerată științifică doar dacă poate fi testată empiric, adică dacă putem colecta date, iar, după analiza acestora, rezultatele obținute ne confirmă sau ne infirmă, după caz, teoria. În această direcție se încadrează și linia urmată de lingvistica bazată pe corpus. LC constituie o metodă de analiză lingvistică și presupune din partea lingvistului *furnizarea dovezilor empirice obținute dintr-un corpus* pentru a-și susține orice teorie lingvistică<sup>303</sup>. O altă cerință științifică urmată în LC este reprezentată de *replicabilitatea rezultatelor*, adică metoda de lucru

<sup>&</sup>lt;sup>299</sup> Paweł Szudarski, op. cit., p. 8.

<sup>&</sup>lt;sup>300</sup> Ibidem.

<sup>&</sup>lt;sup>301</sup> Dan Cristea, Resurse lingvistice și tehnologii ale limbajului natural. Cazul limbii române. Prelegerile Academiei Române. Filiala Iasi, 2005, p. 3.

<sup>&</sup>lt;sup>302</sup> Vaclav Brezina, *Statistics in Corpus Linguistics*. *A practical Guide*, Cambridge University Press, 2018, p. 2.

<sup>&</sup>lt;sup>303</sup> Ibidem.

folosită într-o cercetare trebuie să fie suficient de transparentă astfel încât să poată fi reprodusă în alte studii care îi urmează (în cadrul asa-numitelor *follow-up studies*)<sup>304</sup>.

Corpusul este definit în literatura de specialitate drept o colecție de texte, caracterizată prin: dimensiune mare și posibilitate de prelucrare electronică, realizată cu scopul de a fi reprezentativă pentru limba în cauză și pentru a fi analizată din perspectivă lingvistică (calitativă și cantitativă)<sup>305</sup>. Şi McEnery și Hardie consideră că, atunci când vorbim despre corpus, ne referim la un "un set de texte care, de obicei, sfidează analiza lingvistică cu mâna și ochiul într-un timp rezonabil" [tr. n.]<sup>306</sup>. Până a fi cunoscut așa cum a fost definit mai sus, "termenul corpus se referea la o colecție de texte similare. În secolul al XVIII-lea, cercetătorii făceau referire la un corpus al poeților latini sau la un corpus de legi. Oricum, prima citare a acestui termen se pare că e datată în anul 1956, cu sensul de corp de materiale scrise sau vorbite, care stau la baza unei analize lingvistice" [tr. n.]<sup>307</sup>. Relevanța corpusului lingvistic în cercetare nu a fost niciodată negată, doar perspectivele din care acesta a fost privit fiind diferite, distingându-se două categorii de cercetători.

"Pe de-o parte, unii care consideră că LC este o *metodă* pe care ne putem baza studiile, care ne confirmă sau ne infirmă teorii, iar, pe de altă parte, e cealaltă categorie, cea a cercetătorilor cunoscuți drept *neo-Firthians* (urmași ai lui J. R. Firth), care promovează LC ca pe o *teorie*, iar însuși corpusul lingvistic, în accepțiunea acestora, are statut teoretic"<sup>308</sup>.

Aceștia își bazează afirmația pe ideea conform căreia, folosind LC, obținem noi teorii lingvistice. De pildă, Sinclair susține că "o astfel de abordare evită orice altă clasificare lingvistică anterioară" [tr. n.]<sup>309</sup>. McEnery și Hardie oferă un argument puternic în defavoarea acestei teorii, aducându-l în discuție pe Chomsky, care consideră că informațiile din corpus nu pot fi folosite drept o sursă de înțelegere a naturii limbajului. Astfel, dacă vedem interpretarea grupului *neo-Firthian* prin

<sup>&</sup>lt;sup>304</sup> Vezi și I.S.P. Nation, *op. cit.*, 2016, pp. 118-119. Şi Nation, P., & Crabbe, D., *op. cit.*, 1991, p. 191–201.

<sup>&</sup>lt;sup>305</sup> Michael Stubbs, *Language corpora*, in: Davies, Alan, Catherine Elder, *The Handbook of Applied Linguistics*, Blackwell Publishing Ltd, p. 106 și O'Keefee, Anne, McCarthy, Michael, *From Corpus to Classroom*, Ronald Carter, Cambridge University Press, Cambridge, 2007, p.1-2.

<sup>&</sup>lt;sup>306</sup> Tony McEnery, Andrew Hardie, op. cit., p. 2.

<sup>&</sup>lt;sup>307</sup> După Michael McCarthy, Anne O'Keeffe, *op. cit.*, p. 5.

<sup>&</sup>lt;sup>308</sup> Antonela Ariesan, Lavinia-Iunia Vasiu, op. cit., 2016, p. 955-966.

<sup>&</sup>lt;sup>309</sup> J. Sinclair, *Trust the Text: Language, Corpus and Discourse*, London, Routledge, 2004, p. 191.

raportare la Chomsky, vom ajunge la concluzia că "nimic în afara informațiilor din corpus nu poate fi folosit ca sursă de înțelegere a naturii limbajului" [tr. n.]<sup>310</sup>.

În accepțiunea noastră, LC constituie un mijloc prin care se poate ajunge la un rezultat, o metodă pe care o putem utiliza pentru a ne valida cercetările sau pentru a găsi răspunsuri la unele întrebări complexe, însă nu credem că aceasta poate constitui "teoria" în sine. În plus, întrucât rezultatele ar fi mult prea greu de prelucrat în forma obținută inițial, este absolută nevoie de intervenția statisticii. În acest context, corpusul reprezintă "o mostră de limbă, adică o mică parte dintr-o producție sau dintr-o serie de producții care ne interesează" [tr. n.]<sup>311</sup>, existând puține situații când un corpus poate include toate producțiile lingvistice pe o anumită temă de interes (*the whole population*). Brezina oferă un exemplu în care avem de-a face cu *această situație*, nu doar cu o mostră. Este vorba, bunăoară, despre un corpus format din toate discursurile publice ale lui Margaret Thatcher. Însă e important de reținut că asemenea situații sunt izolate, fiindcă, de cele mai multe ori, nu avem această posibilitate și este nevoie să ne rezumăm la analiza unei mostre de limbă pentru a ne confirma teoria și a ne formula concluziile, bazându-ne, așadar, și pe principiile statisticii<sup>312</sup>.

# 4.1. Corpusul lingvistic. Clasificări

Există mai multe tipuri de corpusuri lingvistice<sup>313</sup>, care de-a lungul timpului au fost clasificate în funcție de diferite criterii. Expunem în continuare clasificările corpusurilor lingvistice așa cum apar la Cristea<sup>314</sup>, conform accepțiunilor lui Sinclair si Ball si Teubert.

- ❖ După *criteriul modalității*, corpusurile se separă în: colecții de texte scrise şi înregistrări de vorbire, care pot fi transcrise<sup>315</sup>.
- În ceea ce privește criteriul explicitării, observăm că avem de-a face cu: corpusuri primare și corpusuri adnotate. Corpusurile primare sunt alcătuite din texte în formatul inițial, destinate uzului uman. În schimb, corpusurile adnotate presupun ca textul primar să fie suplimentat cu adnotări ce reprezintă explicitarea în format inteligibil pentru mașină a informațiilor

<sup>&</sup>lt;sup>310</sup> Tony McEnery, Andrew Hardie, op. cit., p. 148.

<sup>311</sup> Vaclav Brezina, op. cit., p. 6.

<sup>&</sup>lt;sup>312</sup> Vezi, în acest sens, Vaclav Brezina, *op. cit.*, 2018, pp. 6, 15, 16, 18.

<sup>&</sup>lt;sup>313</sup> O clasificare a tipurilor de corpusuri se găsește și la Vaclav Brezina, *op. cit.*, pp. 15-16.

<sup>314</sup> Dan Cristea, Resurse lingvistice și tehnologii ale limbajului natural. Cazul limbii române. Prelegerile Academiei Române. Filiala Iași, 2005, pp. 4-5 (http://profs.info.uaic.ro/~dcristea/papers/cristea-prelegeri.pdf).

- lingvistice și extralingvistice considerate pertinente. Pentru adnotarea corpusurilor se folosesc limbaje specializate<sup>316</sup>.
- ❖ Din perspectiva *cantității*, avem de-a face cu: corpusuri mai mari și corpusuri mai mici. După Cristea, pentru a fi reprezentativ pentru o limbă, un corpus nu poate fi mai mic de 50 de milioane de cuvinte. Un exemplu în acest sens e reprezentat de *British National Corpus*, un corpus de mari dimensiuni (100 de mil. de cuvinte)<sup>317</sup>.
- ❖ Raportarea la *conținut* este extrem de importantă în ceea ce privește corpusurile. Astfel, există: corpusuri de referință și corpusuri speciale. Corpusurile de referință sunt reprezentative pentru o limbă. Ele adună texte care exprimă limbajul scris, dar și vorbit, limbajul formal, dar și informal, și reprezintă categorii socio-intelectuale variate, precum și situații de comunicare diverse. Ponderea limbajului vorbit în corpusurile moderne de referință este de aproximativ 10%. Corpusurile speciale sunt create pentru a satisface un anumit scop.
- ❖ În funcție de *criteriul temporalității*, corpusurile se grupează în: corpusuri pentru o anumită perioadă de timp, corpusuri-monitor și corpusuri atemporale. Primele sunt corpusurile care reprezintă limba caracteristică unei anumite perioade de timp. Corpusurile-monitor înregistrează evoluția limbii în timp, prin conservarea unei colecții de dimensiuni aproximativ constante, care radiografiază în general limba contemporană. În corpusurile atemporale, textele nu sunt selectate după anul apariției.
- ❖ După criteriul criteriul comparabilității, se disting: corpusuri monolingve şi corpusuri multilingve. Corpusurile multilingve sunt folosite în aplicații pentru traducere automată. Printre corpusurile multilingve, se disting corpusurile paralele, care conțin traduceri ale aceluiași text în două sau mai multe limbi<sup>318</sup>.

<sup>316</sup> Conform lui Cristea, cel mai cunoscut este *Extended Mark-Up Language (XML)* https://www.w3.org/TR/REC-xml/, iar ca standarde de adnotare sunt folosite *Corpus Encoding Standard (CES)* https://www.cs.vassar.edu/CES/ și *Text Encoding Initiative (TEI)* https://tei-c.org. *Cf.* Dan Cristea, *op. cit.*, p. 4.

<sup>&</sup>lt;sup>317</sup> Corpus reprezentativ pentru engleza britanică, conținând mostre de limbă scrisă, dar și vorbită: http://www.natcorp.ox.ac.uk

<sup>&</sup>lt;sup>318</sup> După Cristea, un exemplu de corpus multilingv aliniat este *MultextEast*, care conține traducerile aliniate ale *Republicii lui Platon* și ale romanului *1984* de George Orwell, prima în 25 de limbi și a doua în 10 limbi, toate aliniate cu versiunea engleză, cf. Cristea, *op. cit.*, pp. 4-5.

Dintre acestea, se distinge un tip special de corpus, și anume *corpusul vorbitorilor nonnativi* (*learner corpus*), care poate conține atât colecții de texte scrise, cât și producții orale.

În ceea ce privește limba română ca limbă maternă (RLM), apreciem în mod deosebit faptul că s-au elaborat sau se află în dezvoltare diverse corpusuri, în diferite centre universitare, precum și la Academia Română, prin Institutul de Cercetări pentru Inteligența Artificială din București și Institutul de Informatică Teoretică din Iași. Prezentăm, în continuare, succint, principalele corpusuri pentru RLM, începând cu cele electronice<sup>319</sup>:

Corpusul *CoRoLa* (*Corpus Computațional de Referință pentru limba română ca limbă contemporană*) este un proiect al Academiei Române, în derulare la Institutul de Cercetări pentru Inteligența Artificială (ICIA) și Institutul de Informatică Teoretică – Iași (IIT)<sup>320</sup>. Corpusul conține texte diverse, începând cu anul 1989 până în prezent. Este alcătuit din peste un miliard de cuvinte și 300 de ore de înregistrări, din 70 de domenii științifice<sup>321</sup>. Textele sunt "adnotate cu metainformații – precum autor, data publicării etc. – și cu date lingvistice – precum părți de vorbire, forma din dicționar a cuvântului adnotat, dependențe sintactice etc."<sup>322</sup>.

Corpusul *ROMBAC* (*The Romanian Balanced Annotated Corpus*)<sup>323</sup> este realizat la Institutul pentru Inteligență Artificială al Academiei Române și reprezintă cel mai mare corpus adnotat pentru limba română. ROMBAC conține cinci subcorpusuri egale (de aproximativ 7 milioane de cuvinte), din genurile: jurnalistic, juridic, ficțiune, medicină și farmacie, biografii și lucrări de critică literară. Întregul corpus conține în jur de 36 milioane de cuvinte.

Corpusul *RoCo\_News* este un corpus specializat, conținând aproximativ 7 milioane de unități lexicale (partea de texte jurnalistice din corpusul ROMBAC).

Dan Tufis, Dan Cristea, *An Insight into CoRoLa*, prezentare invitată la SpeD 2017 (https://profs.info.uaic.ro/~dcristea/Talks/SPeD2017%20talk\_DT\_DC.ppt.pdf).

<sup>&</sup>lt;sup>319</sup> O prezentare a corpusurilor existente pentru RLM a făcut și Lavinia Vasiu, în cap. *4.2. Corpusurile de limba română*, în *op. cit.* 

<sup>&</sup>lt;sup>320</sup> În acest sens, vezi CoRoLa, disponibil online: http://corola.racai.ro.

<sup>&</sup>lt;sup>322</sup> Elena Irimia, "Accelerarea dezoltării unui corpus digital adnotat cu relații de dependență pentru limba română utilizând resurse și instrumente construite pentru alte limbi", în *Revista Română de Informatică* și *Automatică*, vol. 25, nr. 3, 2015, p. 6.

<sup>&</sup>lt;sup>323</sup> Descris în: Radu Ion, Elena Irimia, Dan Ștefănescu, Dan Tufiș, *ROMBAC: The Romanian Balanced Annotated Corpus*, în *LREC*, 2012, pp. 339-344.

Așadar, și acest corpus a fost elaborat la Institutul pentru Inteligență Artificială al Academiei Române<sup>324</sup>.

Pe lângă aceste corpusuri, s-a elaborat și un număr semnificativ de corpusuri în format imprimat. Reținem aici doar câteva exemple<sup>325</sup>: *Corpus de română vorbită* (CORV)<sup>326</sup>, Româna vorbită actuală (ROVA)<sup>327</sup>, Interacțiunea verbală în limba română actuală (IVLRA)<sup>328</sup>, Interacțiunea verbală (IV II)<sup>329</sup>, Corpus de limbă română vorbită actuală<sup>330</sup>. Se observă că o trăsătură a acestora constă în faptul că "locutorii înregistrați au un grad peste mediu de educație, iar limba, caracterizată de ceea ce gramaticile normative numesc greșeli, este puțin reprezentată"<sup>331</sup>, așadar, nu pot fi folosite în domeniul achiziției L2, unde este mare nevoie de evidența acestor erori pentru a putea compara producțiile vorbitorilor nonnativi cu cele ale locutorilor nativi.

Prin urmare, întrucât pentru limba română ca limbă străină există doar de curând un corpus publicat de producții ale nonnativilor (*Learner Corpus*)<sup>332</sup>, am considerat necesară elaborarea unui material conținând producțiile scrise ale studenților din programele de studiu An pregătitor și Erasmus, de la Universitatea Babeș-Bolyai din Cluj-Napoca<sup>333</sup>, corpus care va fi prezentat în capitolul 6.

<sup>&</sup>lt;sup>324</sup> Dan Tufiş, Elena Irimia, "RoCo-News: A Hand Validated Journalistic Corpus of Romanian", în *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, 2006, pp. 869-872.

<sup>&</sup>lt;sup>325</sup> Pentru o prezentare mai detaliată a corpusurilor imprimate, disponibile pentru limba română, vezi Carmen Mîrzea Vasile, "Corpusurile lingvistice de limba română şi importanța lor în realizarea de materiale didactice pentru limba română ca limbă străină" (lucrare prezentată la Conferința aniversară 40 de ani de limba română ca limbă străină la UBB.1974-2014, Facultatea de Litere, UBB, Cluj-Napoca, 3-4 octombrie 2014).

<sup>&</sup>lt;sup>326</sup> Laurenția Dascălu Jinga, *Corpus de română vorbită (CORV). Eșantioane*, București, Oscar Print, 2002.

<sup>327</sup> Laurenția Dascălu Jinga (coord.), Româna vorbită actuală (ROVA). Corpus şi studii, Academia Română, Institutul de Lingvistică "Iorgu Iordan – Al. Rosetti", Bucureşti, Editura Academiei Române, 2011.

<sup>&</sup>lt;sup>328</sup> Liliana Ionescu-Ruxăndoiu (coord.), *Interacțiunea verbală în limba română actuală. Corpus (selectiv). Schiță de tipologie*, București, Editura Universității din București, 2002.

<sup>&</sup>lt;sup>329</sup> Liliana Ionescu-Ruxăndoiu (coord.), *Interacțiunea verbală (IV II). Aspecte teoretice și aplicative. Corpus*, București, Editura Universității din București, 2007.

<sup>&</sup>lt;sup>330</sup> Luminița Hoarță Cărăușu (coord.), *Corpus de limbă română vorbită actuală*, Iași, Editura Tehnică, Științifică și Didactică Cermi, 2005 și Hoarță Cărăușu, Luminița (coord.), *Corpus de limbă română vorbită actuală*, Iași, Editura Universității "Alexandru Ioan Cuza", 2013.

<sup>&</sup>lt;sup>331</sup> Carmen Mîrzea Vasile, op. cit.

<sup>&</sup>lt;sup>332</sup> Este vorba despre: Mihaela-Viorica Constantinescu, Gabriela Stoica, *Româna ca limbă străină*. *Corpus*, Editura Universității din București, București, 2020, descris în secvența următoare.

<sup>333</sup> În ceea ce priveşte producțiile orale, există deja, începând cu anul 2017, CORLS, întocmit de Lavinia-Iunia Vasiu.

## 4.2. Avantajele unui corpus de producții scrise ale vorbitorilor nonnativi

Corpusul de producții scrise ale vorbitorilor nonnativi este definit drept "o colecție de producții scrise și orale ale vorbitorilor nonnativi" [tr. n.]<sup>334</sup>. Acesta e utilizat, tot mai frecvent, în procesele de predare, învățare și evaluare ale unei limbi străine, atât în cadrul metodei cunoscute în literatura străină de specialitate sub denumirea de *data-driven learning (DDL)*, în accepțiunea urmașilor lui Firth, cât și ca metodă empirică de verificare și confirmare a variatelor teorii. Modificări semnificative se pot observa în felul în care este percepută limba, datorită existenței și utilizării unui corpus al vorbitorilor nonnativi: "se mută accentul de pe dominația vorbitorului nativ și se aduce *limba* nonnativului în centrul atenției, oferind posibilitatea, la nivelul clasei, să se folosească o *limbă* pe care o pot crea și pe care o pot folosi cei care o învață" [tr. n.]<sup>335</sup>. Așadar, se pliază mult mai bine pe nevoile grupului-țintă în discuție.

Un exemplu în acest sens vine din studiile recente asupra inputului la care sunt expuşi vorbitorii nonnativi. La Ferguson se întâlneşte expresia *vorbire pentru străini* (*foreign talk*) pentru a desemna acest tip de input, care se caracterizează prin faptul că se aseamănă cu limbile pidgin<sup>336</sup>, adică, "se pot observa omisiuni (ale flectivelor, ale cuvintelor funcționale), adăugiri redundante (pronume, de pildă, pentru facilitarea înțelegerii: M-a văzut PE MINE.), reordonarea constituienților enunțului etc."<sup>337</sup>. Pe teren românesc, Platon propune *microlimba* (ML) drept termen corespondent. Astfel, din dorința de "a asigura succesul comunicării exolingve cu deținătorul unui nivel scăzut de interlimbă (IL), locutorul nativ este nevoit să facă anumite concesii, «coborând» nivelul LN cât mai aproape de capacitatea de înțelegere a interlocutorului său" [tr. n.]<sup>338</sup>. În acest fel, vorbitorul nativ "ajunge să vehiculeze o limbă mult simplificată în comparație cu cea utilizată în interacțiunile curente cu vorbitorii nativi, pe care noi am numit-o "microlimbă" (ML)" [tr.n.]<sup>339</sup>. Ce e de remarcat la ML e că, deși vorbim despre o limbă simplificată și adaptată în funcție de IL vorbitorului nonnativ, aceasta se caracterizează drept "*corectă* din

<sup>334</sup> Michael McCarthy, Anne O'Keeffe, op.cit., p. 8.

<sup>&</sup>lt;sup>335</sup> *Ibidem*.

<sup>&</sup>lt;sup>336</sup> C. Ferguson, "Absence of copula and the notion of simplicity: a sudy of normal speech, baby talk, foreigner talk and pidgins", în D. Hymes (ed.), *Pidginization and Creolization of Languages*, Cambridge, Cambridge University Press, 1971, *apud* Lavinia Vasiu, *op. cit.*, 2020, p. 83.

<sup>&</sup>lt;sup>337</sup> Lavinia Vasiu, op. cit., 2020, p. 83.

<sup>&</sup>lt;sup>338</sup> Elena Platon, "The "micro-language" – an assault on natural languages?", în *Lingua. Language and culture*, XVIII, nr. 2, 2019a, p. 68.
<sup>339</sup> *Ibidem*.

punct de vedere gramatical, însă mult mai săracă" [tr.n.]<sup>340</sup>. Pe de altă parte, interlimba este diferită de la un individ la altul și este definită drept "un sistem lingvistic structurat, care functionează după reguli/principii proprii ce nu apartin pe deplin nici L1, nici L2, și care se află într-un continuu proces de restructurare în scopul apropierii de sistemul L2" [tr.n.]<sup>341</sup> sau ca un ..sistem lingvistic hibrid. insuficient stabilizat, aflat undeva pe drumul dintre L1 și L2"342 ce îl caracterizează explusiv pe cursant. În literatura de specialitate, conceptul de IL a fost vehiculat de către Selinker<sup>343</sup>, Corder<sup>344</sup>, iar pe plan national, importanța acestui concept a fost semnalată, mai întâi, în 2015, de către Platon<sup>345</sup>, iar apoi, de către Vasiu<sup>346</sup>. Pentru cercetarea noastră, conceptul de *microlimbă* este relevant, întrucât inventarul lexical al RLS, realizat pe baza corpusului de producții scrise ale vorbitorilor nonnativi, ar constitui un instrument în stabilirea microlimbii pentru fiecare nivel, în scopul obținerii unității în ceea ce privește inputul la care sunt expuși nonnativii. Principalul aspect care i s-ar putea imputa acestei ML este că nu se mai pot folosi texte autentice, având în vedere că limba este "adaptată" nivelului interlocutorului nonnativ, situație care poate fi solutionată prin adoptarea unei "reprezentări lărgite asupra notiunii de autenticitate"347, optându-se pentru o autenticitate "conferită de tematica aleasă, de corespondenta inputului cu situația reală de comunicare, de maniera în care formulăm sarcinile"348, aspecte ce îi creează cursantului un cadru similar celui din "viata reală".

După cum am putut constata, corpusul vorbitorilor nonnativi reprezintă un pas înainte, de dată relativ recentă, dar foarte important în domeniul CL. Granger definește corpusul vorbitorilor nonnativi drept "o colecție de texte *autentice*, produse de vorbitori străini sau de minorități. Aceasta arată că editurile, dar și universitățile din străinătate au început să colecteze și să analizeze producții ale vorbitorilor

<sup>&</sup>lt;sup>340</sup> Elena Platon, *op. cit.*, 2019a, p. 69.

<sup>&</sup>lt;sup>341</sup> *Ibidem*, p. 68.

<sup>&</sup>lt;sup>342</sup> Elena Platon, *Româna ca limbă străină (RLS). Elemente de metadidactică*, Editura Presa Universitară Clujeană, Cluj-Napoca, 2021, p. 344.

<sup>&</sup>lt;sup>343</sup> L. Selinker, *Language Transfer*, în "General Linguistics", 9/2, 1969, pp. 67-92 apud Elena Platon, op. cit. 2021, pp. 343-344.

<sup>&</sup>lt;sup>344</sup> S.P. Corder, *La sollicitation de données d'interlangue*, în "Langages", *nr. 57*, 1980, p. 29 *apud* Elena Platon, *op. cit.*, 2021, p. 344.

<sup>&</sup>lt;sup>345</sup> Elena Platon, "Reflexii asupra conceptului de *interlimbă*", în vol. *Înspre și dinspre Cluj. Contribuții lingvistice. Omagiu profesorului G.G. Neamțu la 70 de ani, Editura Argonaut și Editura Scriptor*, Cluj-Napoca, 2015, pp. 527-539.

<sup>&</sup>lt;sup>346</sup> Lavinia-Iunia Vasiu, op. cit., 2020.

<sup>&</sup>lt;sup>347</sup> Elena Platon, *op. cit.*, 2021, p. 367 și David Atkinson, *Teaching in the target language: a problem in the current orthodoxy*, în "The Language Learning Journal", 8:1, 1993, pp. 2-5, DOI: 10.1080/09571739385200261.

<sup>&</sup>lt;sup>348</sup> Elena Platon, op. cit., 2020, p. 368.

nonnativi pentru a produce materiale didactice bazate pe corpus." [tr. n.]<sup>349</sup>. Mostrele de limbă scrisă sunt permanente, ceea ce le face să fie și mai simplu de colectat. Acesta este și motivul pentru care în lume, în ultimii ani, s-au demarat mai multe proiecte care au avut drept obiectiv elaborarea unor corpusuri de producții scrise ale vorbitorilor nonnativi. Un alt motiv pentru care aceste materiale au intrat în sfera de interes a cercetătorilor se datorează și disponibilității a tot mai multor aplicații de prelucrare a corpusului.

Unul dintre cele mai cunoscute astfel de proiecte este *International Corpus* of Learner English (ICLE)<sup>350</sup>. Acesta este un corpus computerizat, alcătuit din eseuri argumentative, pe diferite teme, redactate de studenți la limba engleză, aflați la nivel avansat, în anul doi sau trei de studiu. El conține o serie de subcorpusuri, organizate în funcție de limba maternă (L1) a studenților<sup>351</sup>. Ellis și Barkhuizen arată că "cele mai multe corpusuri ale vorbitorilor nonnativi sunt realizate din materiale scrise de studenți avansați, iar, pentru a fi utile cercetării, ar trebui să cuprindă mai multe paliere, inclusiv limba vorbită" [tr. n.]<sup>352</sup>. Autorii atrag atenția, așadar, asupra necesității utilizării unor date cât mai variate pentru ca aceste corpusuri să poată fi cu adevărat reprezentative pentru limba în discuție: producții pe niveluri diferite, texte aparținând unor stiluri diferite, din contexte cât mai diverse etc.

Corpusurile vorbitorilor nonnativi sunt, la rândul lor, clasificate în funcție de o serie de criterii<sup>353</sup>. După mijlocul de comunicare, avem corpusuri cu texte scrise și corpusuri de limbă vorbită. Un alt criteriu este reprezentat de limbă, atât L1, cât și L2, și, din acest punct de vedere, corpusurile pot fi monolingve sau bilingve. Un aspect de reținut în cazul elaborării corpusurilor monolingve este acela ca reprezentarea L1 să fie cât mai echilibrată în cadrul corpusului, afirmație combătută de către Vasiu, care nu consideră că "o distribuție inegală a L1 într-un corpus ar trebui să afecteze puternic rezultatele analizei"<sup>354</sup>. Mai mult, autoarea susține tocmai nevoia unui număr semnificativ de producții ale studenților cu aceeași L1 "pentru a permite extragerea trăsăturilor specifice L1"<sup>355</sup>.

\_

<sup>&</sup>lt;sup>349</sup> Granger *apud* Anne O'Keefee, Michael McCarthy, *From Corpus to Classroom*, Ronald Carter, Cambride University Press, Cambridge, 2007, p.24.

<sup>&</sup>lt;sup>350</sup> Elaborat la Universitatea Catolica din Louvin, de către un colectiv condus de către Sylvienne Granger.

<sup>&</sup>lt;sup>351</sup> Rod Ellis, Gary Barkhuizen, *Analysing Learner Language*, Oxford University Press, 2005, p. 29.

<sup>&</sup>lt;sup>352</sup> *Ibidem*, p. 357.

<sup>&</sup>lt;sup>353</sup> După Gaëtanelle Gilquin, "From design to collection of learner corpora", în S. Granger, G. Gilquin, F. Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*, 2015, pp. 12-22.

<sup>354</sup> Lavinia-Iunia Vasiu, op. cit., p. 125.

<sup>355</sup> Ibidem.

Corpusurile sunt clasificate și în funcție de momentul la care au fost incluse producțiile. Se disting, astfel, trei tipuri de corpusuri: diacronice/longitudinale (care urmăresc parcursul unui vorbitor în procesul de achiziție a L2), semilongitudinale (producții realizate în diferite momente de învățare a limbii, însă de la vorbitori diferiți) și sincronice (semnificative pentru un anumit moment al parcursului de achiziție a L2). Desigur că, pe lângă aceste aspecte, sunt importante cadrul și timpul avut la dispoziție de subiect pentru a produce discursul, precum și variabilele generale (sexul, nivelul de educație etc.), dar și variabilele specifice/individuale (L1 și alte L2, vârsta, nivelul de instruire, motivație etc.).

În ciuda faptului că, pentru limba română ca limbă maternă, există diferite corpusuri, electronice sau tipărite<sup>356</sup>, în domeniul RLS, situația nu este la fel de mulțumitoare în ceea ce privește acest aspect. Totuși, primii pași au fost făcuți odată cu apariția Corpusului Oral de Limba Română ca Limbă Străină (CORLS), elaborat, în 2017, de către Vasiu, și a volumului *Româna ca limbă străină. Corpus*, realizat, în 2020, de către Constantinescu și Stoica<sup>357</sup>. CORLS conține producțiile orale a 172 de studenti din Anul pregătitor (de la Departamentul de limbă, cultură și civilizatie românească al Facultății de Litere din cadrul Universității Babes-Bolyai din Cluj-Napoca), înregistrate în timpul examinărilor orale de nivelul A1, de-a lungul a trei ani universitari, în perioada 2014-2017<sup>358</sup>. Pe de altă parte, *Româna ca limbă străină*. Corpus "include un corpus de texte autentice de română ca limbă străină (RLS), produse de vorbitori nonnativi, în cadrul programului de An Pregătitor al Universității din București"359. Acest din urmă corpus contine atât producții scrise (380 de texte), colectate din producțiile studenților de la examene, în cadrul activităților la clasă și din temele de casă<sup>360</sup>, cât și transcrieri ale unor producții orale (79 de producții orale)<sup>361</sup>. Cu toate acestea, necesitatea și utilitatea unor astfel de materiale au fost deja semnalate:

"Corpusurile lingvistice reprezentative sunt foarte importante în descrierea corectă a unei limbi. Reprezentând o mostră de limbă standard, neutră (varianta de limbă pe care o învață nenativii în mod ideal, abstract), utilizarea unui asemenea

56

<sup>&</sup>lt;sup>356</sup>Vezi *supra* 4.1.

<sup>&</sup>lt;sup>357</sup> Vezi Mihaela-Viorica Constantinescu, Gabriela Stoica, op. cit. 2020.

<sup>358</sup> Lavinia-Iunia Vasiu, op. cit., 2020.

<sup>&</sup>lt;sup>359</sup>După prezentarea volumului, pe pagina Editurii Univeristății din București: https://editura-unibuc.ro/produs/romana-ca-limba-straina-corpus/.

<sup>&</sup>lt;sup>360</sup> Pentru o descriere a corpusului scris, vezi Constantinescu, Stoica, op. cit., 2020, pp. 24-28.

<sup>&</sup>lt;sup>361</sup> Pentru detalii legate de corpusul oral, vezi Constantinescu, Stoica, op. cit., 2020, pp. 28-31.

corpus este absolut necesară și în realizarea de materiale didactice pentru nenativi"<sup>362</sup>.

Dintre toate avantajele pe care un astfel de instrument le poate avea, se remarcă, asadar, acela de a sta la baza elaborării unor materiale didactice reprezentative, atât în ceea ce priveste procesul de predare (manuale, auxiliare, dictionare, curricule, pe nivelurile corespunzătoare CECRL), cât și cel de evaluare (teste și grile de evaluare). Dacă româna ca limbă străină ar beneficia de un astfel de corpus, corelatia dintre ceea ce i se cere studentului si ceea ce poate realiza, de fapt, va fi mult mai aproape de realitate. Cel mai elocvent exemplu în acest sens considerăm că este tot proiectul English Profile din care au rezultat English Vocabulary Profile și English Grammar Profile. Aceste instrumente de cercetare oferă o inventariere clară pe nivelurile de referință din CECRL a elementelor de vocabular utilizate de studenti la fiecare palier, împreună cu exemple din productiile acestora, ceea ce asigură o garanție în plus a echilibrului dintre obiectivele propuse în predare și cele din evaluare. Un material similar ar fi extrem de util și limbii române, însă acesta nu se poate realiza fără un corpus al vorbitorilor nonnativi. Pe de altă parte, în English Grammar Profile structurile gramaticale sunt corelate, prin analiza corpusului, cu ceea ce este capabil efectiv studentul să utilizeze, fără a se apela, la experientă, la intuiție și la comparația cu alte limbi, ceea ce în cazul RLS încă se întâmplă frecvent. Prin urmare, existând un astfel de material și pentru limba română, s-ar putea asigura o uniformizare atât la nivelul unei singure instituții, cât și al unuia mai larg, national si european. Din păcate, în lipsa corpusului si a materialelor rezultate pe baza acestuia, fiecare cadru didactic, respectiv, fiecare organizatie de predare si de evaluare se raportează, în cel mai bun caz, la curriculumul din propriul centru, sau la aspectele care sunt deja descrise<sup>363</sup>, deși, de cele mai multe ori, intuitia pare să rămână elementul hotărâtor<sup>364</sup>.

Încă o dată se dovedește importanța principiului care a stat la baza realizării proiectului *English Profile Wordlists*: e mai important să se inventarieze limba pe care o știu și o folosesc vorbitorii din toată lumea decât cea vorbită de nativi. Avantajul utilizării unui astfel de corpus, conținând producții atât de diferite, și

<sup>&</sup>lt;sup>362</sup> Carmen Mîrzea Vasile, op. cit.

<sup>&</sup>lt;sup>363</sup> Ne referim aici la manualele deja existente, realizate în același mod, sau la: Victoria Moldovan, Liana Pop, Lucia Uricaru, *op. cit.*, 2001; Elena Platon, Ioana Sonea, Lavinia Vasiu, Dina Vîlcu, *op. cit.*, 2014.

<sup>&</sup>lt;sup>364</sup> Vezi, în acest sens, Antonela Arieşan, Lavinia-Iunia Vasiu, "On the Necessity of a Learner Corpus – Romanian as a Foreign Language (RFL)", în *Globalization and National Identity. Studies on the Strategies of Intercultural Dialogue*, ed. Iulian Boldea, Arhipelag XXI, Târgu Mureş, 2016, p. 955-966.

scrise, și orale, realizate de către persoane cu profiluri așa de diverse, este acela că ne poate oferi o descriere mult mai precisă a conținuturilor pe niveluri. "De asemenea, faptul că datele din corpus provin de la vorbitori din toată lumea, face ca rezultatele obținute:

- să nu fie concentrate numai asupra limbii vorbite de nativi;
- să nu fie relevante numai pentru anumiți vorbitori, din anumite medii, întrucât limbile materne ale acestor vorbitori sunt atât de diverse cum sunt și vârstele și preocupările acestora" [tr. n.]<sup>365</sup>.

Biber susține că, în cazul unui corpus, mai important este grupul-țintă pentru care acesta e întocmit, respectiv publicul căruia îi este adresată lista de cuvinte elaborată, și nu atât de mult dacă avem mostrele cele mai bune de limbă<sup>366</sup>. Nation susține și el această afirmație, însă evidențiază că, în cazul achiziției L2, este foarte greu de descris grupul-țintă<sup>367</sup>. Totuși, dacă ne referim la cazul nostru, în special, putem obține o caracterizare destul de bună a publicului, printr-o analiză a studenților din ultimii ani, întrucât vorbim despre studenți care provin din aproximativ aceleași țări, cu aceleași limbi, materne și străine, cunoscute. Din analiza datelor biografice ale studenților din perioada 2013-2016<sup>368</sup>, respectiv 2013-2018<sup>369</sup>, observăm preponderența studenților vorbitori de limbă arabă, aproximativ 2/3 din totalul populației analizate), urmați de cei vorbitori de albaneză, limbi asiatice și franceză, conform graficelor de mai jos:

<sup>365</sup> Harrison, Julia, Barker, Fiona (eds.), op. cit., p. 4.

<sup>&</sup>lt;sup>366</sup> Biber, D., "Representativeness in corpus design", în *Literary and Linguistic Computing*, 8(4), 1993, pp. 243–257.

<sup>&</sup>lt;sup>367</sup> I.S.P. Nation, *op. cit.*, 2016, p. 95.

<sup>&</sup>lt;sup>368</sup> Dina Vîlcu, Antonela Arieşan, Lavinia-Iunia Vasiu, *The use of test takers productions in redesigning writing assessment grids: a corpus based study* (lucrare prezentată la *ALTE* 6th International Conference, 3-5 mai 2017, Bologna, Italia).

<sup>&</sup>lt;sup>369</sup> Dina Vîlcu, Antonela Arieşan, *Proposal for illustration of the supplement with a test for end of LSP course (medical profile)*, la atelierul "Limbajul pentru scopuri specifice - Language for Specific Purposes (LSP)", (lucrare prezentată în cadrul întâlnirii bianuale ALTE Multilingvismul și utilizarea limbajului pentru scopuri specific – Rolul evaluării, 11-13 aprilie 2017, Cluj-Napoca).

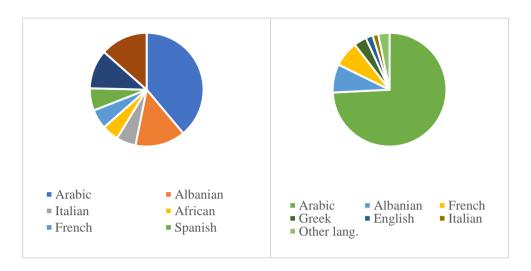


Figura 4 - L1 în cazul studenților din an pregătitor, în perioada 2013-2016 (stânga), respectiv, 2013-2018 (dreapta)

O altă observație importantă legată de corpus este necesitatea relaționării cu vârsta celor din grupul-țintă. Astfel, Nation arată că, dacă avem de-a face cu o listă pentru copii, de exemplu, avem nevoie de un corpus care reflectă limba copiilor, mai mult, pentru a fi cu adevărat relevant, acesta trebuie să păstreze erorile pe care le-au făcut copiii în producții<sup>370</sup>. Desigur că aceeași recomandare va fi aplicabilă și în situația noastră. Având în vedere că studenții ale căror producții le-am preluat în corpusul nostru se găsesc în plin proces de achiziție a limbii române, considerăm că erorile făcute de către aceștia în producțiile lor scrise sunt important de semnalat, deoarece ne pot releva unele elemente poate inadecvate pentru un anumit nivel sau orice alte inadvertențe.

În ceea ce privește dimensiunea corpusului, opiniile sunt împărțite. Sorell a arătat că un corpus de aproximativ 20 de milioane de cuvinte ar fi necesar pentru a obține liste relevante, de aproximativ 9000 de cuvinte<sup>371</sup>. Aceeași valoare o susțin și

<sup>&</sup>lt;sup>370</sup> I.S.P. Nation, op. cit., 2016, p. 98-99.

<sup>&</sup>lt;sup>371</sup> C. J. Sorell, *A study of issues and techniques for creating core vocabulary lists for English as an international language*. Unpublished PhD thesis, Victoria University of Wellington, New Zealand, 2013, p. 204, *apud* I. S.P. Nation, *op. cit.*, 2016, p. 99.

Brysbaert și New<sup>372</sup>, în vreme ce Brezina și Gablasova<sup>373</sup> consideră că este relevant chiar și un corpus de un milion de cuvinte sau de 100 de milioane sau chiar și mai mare. Nation arată că, în cazul cuvintelor cu frecvență ridicată, avem nevoie de un corpus de 20 de milioane de cuvinte dacă dorim să avem cel puțin 50 de ocurențe ale fiecărui cuvânt astfel încât să putem obține suficiente exemple necesare<sup>374</sup>.

După Nation, cel mai comun motiv pentru care se elaborează o listă de cuvinte este acela de a putea decide, în cunoștință de cauză, ce elemente de vocabular trebuie să învețe studenții de L2 în primii ani de studiu. Mai mult, dacă ne raportăm și la vârsta studenților, va fi și mai greu de obținut o listă relevantă. De altfel, Nation este de părere că "pare imposibil să elaborăm o listă generală, care să conțină mai mult decât primele 1000 de cuvinte și care să fie relevantă pentru studenții de toate vârstele și cu nevoi de învățare diferite" [tr. n.]<sup>375</sup>. Așadar, încă o dată, se arată cât de important este grupul-țintă căruia îi e adresată o asemenea listă.

Listele de cuvinte au o utilitate deosebită mai ales în ceea ce privește evaluarea. Secțiunea de vocabular a testelor *LATOS* (*The Language Achievement Test for Overseas Students*) s-a bazat pe o listă dezvoltată de către Campion și Elley în Noua Zeelandă, obținută dintr-un corpus de doar 300.000 de cuvinte<sup>376</sup>. Un alt exemplu de teste bazate pe liste este reprezentat de *Vocabulary Size Tests*, elaborate de către Nation și Beglar<sup>377</sup> și Coxhead, Nation și Sim<sup>378</sup>, pornind de la BNC/COCA lists.

Corpusul își găsește utilitatea chiar și în procesul de pregătire a introducerii unor cuvinte, deoarece va ajuta la asigurarea unui control, dar și a unei progresii în predarea și, mai apoi, în testarea diferitelor sensuri ale unui element lexical. Astfel, se susține ideea conform căreia, folosind numai frecvența cuvintelor în textele scrise de vorbitorii nativi, nu vom ajunge întotdeauna la rezultate relevante. Tocmai de aceea, în urma calculului frecvenței, rezultatele trebuie trecute și printr-un corpus al producțiilor vorbitorilor nonnativi. Nici utilitatea unui corpus al vorbitorilor nativi, pe lângă cel al vorbitorilor nonnativi, în cadrul orelor de RLS, nu este însă negată,

<sup>&</sup>lt;sup>372</sup> M. Brysbaert, B. New, "Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior", în *Research Methods*, 41(4), 2009, pp. 977–990.

<sup>&</sup>lt;sup>373</sup> V. Brezina, D. Gablasova, "Is there a core general vocabulary? Introducing the New General Service List.", în *Applied Linguistics*, 36(1), 2015, pp. 1–22.

<sup>&</sup>lt;sup>374</sup> I.S.P. Nation, op. cit., 2016, p. 100.

<sup>&</sup>lt;sup>375</sup> *Ibidem*, p. 113.

<sup>&</sup>lt;sup>376</sup> *Ibidem*, p. 118.

<sup>&</sup>lt;sup>377</sup> P. Nation, D. Beglar, "A vocabulary size test", în *The Language Teacher*, 31(7), 2007, pp. 9–13.

<sup>&</sup>lt;sup>378</sup> A. Coxhead, P. Nation, D. Sim, "Creating and trialling six versions of the Vocabulary Size Test", în *The TESOLANZ Journal*, 22, 2014, 13–27.

întrucât profesorii care predau L2 ar putea fi interesați de volumul vocabularului vorbitorilor nativi. O asemenea informație poate oferi unele indicații asupra sarcinilor pe care ar trebui să le propună spre realizare, mai ales că vorbitorii de română ca limbă străină vor studia sau vor lucra alături de nativi<sup>379</sup>.

Cu toate că vorbim despre un corpus al vorbitorilor nonnativi, considerăm că acesta își poate dovedi aplicabilitatea și în domeniul L1. Astfel, el poate constitui o sursă pentru exemple relevante, din limba scrisă sau vorbită, autentice, și nu construite artificial sau preluate din texte literare. În plus, poate reprezenta o bază pentru cercetări științifice viitoare, în toate domeniile lingvisticii – punctuație, ortografie, fonetică, morfologie, sintaxă, pragmatică, semantică etc.

Indiferent de metoda aleasă, cercetătorii anglofoni au concluzionat că este, evident, nevoie ca elevul să cunoască primele aproximativ 3000 dintre cele mai frecvente cuvinte din limbă. Acestea sunt considerate a fi prioritare și nu se recomandă învățarea altora înainte de a le cunoaște pe acestea<sup>380</sup>. Mai mult, Nation susține că, "după ce aceste 3000 de cuvinte au fost însușite, este mai important să ne concentrăm asupra dezvoltării unor strategii de înțelegere și învățare a cuvintelor cu frecvență mai redusă în limbă" [tr. n.]<sup>381</sup>. Pe de altă parte, în susținerea acestei afirmații, vine și teoria lui Sinclair și Renouf, teorie legată de *lexical syllabus*, adică posibilitatea de a obține, din corpus, o listă de cuvinte necesară profesorului în predare. Însă, conform autorilor, aceasta nu este nicidecum o simplă listă de cuvinte, ci una care reține diferitele aspecte ale limbii, bazată pe criteriul frecvenței. Acești autori consideră că "accentul ar trebui să cadă asupra a trei aspecte:

- cele mai cunoscute forme ale cuvintelor în limbă;
- tiparul de utilizare;
- combinațiile pe care acele cuvinte le formează" [tr. n.]<sup>382</sup>.

Ideea din spatele acestei afirmații este aceea că cele mai frecvente cuvinte au și o mulțime de sensuri. Așadar, nu este nevoie să se învețe foarte multe cuvinte, fiind importante, mai ales, acele cuvinte care pot desemna cât mai multe idei. Un exemplu în acest sens, în cazul limbii române, îl poate reprezenta verbul *a face*, cel mai prolific verb din limba română, care intră în combinații diferite, concrete (*a face* 

<sup>&</sup>lt;sup>379</sup> Paul Nation and Robert Waring, *Vocabulary Size, Text Coverage and Word Lists*, în N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, Acquisition, and Pedagogy*, Cambridge University Press, Cambridge, 1997, pp. 6-19.

<sup>&</sup>lt;sup>380</sup> Doar prin comparație, deocamdată, putem presupune și noi că această observație s-ar putea aplica și limbii noastre.

<sup>&</sup>lt;sup>381</sup> Paul Nation and Robert Waring, op. cit., 1997, p. 11.

<sup>&</sup>lt;sup>382</sup> Tony McEnery, Course in *Corpus Linguistics: Method, Analysis, Interpretation* (curs online pe platforma *futurelearn.com*), Lancaster University.

o plimbare/curățenie/mâncare/tema/o casă etc.) sau mai abstracte (a face descoperiri/a face aranjamentele necesare etc.).

Pe lângă toate aceste argumente privitoare la materialele de predare și evaluare, "un avantaj esențial pe care acest tip de corpus îl poate aduce în clasă este *autenticitatea*" [tr. n.]<sup>383</sup>, deoarece se vor putea prelua exemple din corpus pentru a se ilustra contextele de utilizare ale cuvintelor sau ale structurilor gramaticale nou introduse, exemple care, deocamdată, se oferă tot pe baza intuiției de vorbitor nativ sau, la nivelurile inferioare (A1-A2), acestea sunt construite artificial.

Corpusul în sine, din perspectiva DDL, poate avea utilitate imediată în cadrul orelor de RLS atât pentru profesori, cât si pentru studenti, în extragerea regulilor gramaticale sau de descoperire a diferitelor relatii dintre anumite structuri avute în vedere în corpus. Gilquin și Granger arată că, astfel, "studenții devin mai implicati, mai activi și, în final, autonomi în procesul de învățare" [tr. n.]<sup>384</sup>. În continuare, autoarele arată că, utilizând astfel corpusul, studentul va câstiga mai multă încredere în sine, ceea ce nu poate decât să ajute la însusirea cât mai eficientă a limbii. În plus, considerăm noi, studentul se va obisnui să lucreze cu un material de mari dimensiuni, ceea ce în cazul studenților străini din anul pregătitor de limbă română poate constitui o adevărată provocare, întrucât mulți dintre ei, atunci când intră la cursurile de română, nu sunt obișnuiți să prelucreze texte, nici măcar în limbile lor. Având un corpus al vorbitorilor nonnativi, profesorul, încă de la nivelul A1, ar putea aplica metoda prin descoperire. De pildă, identificarea diferentelor dintre genuri, la substantiv, identificarea formelor verbale, a adjectivelor sau a unor familii de cuvinte etc. s-ar putea învăța pe baza diferitelor exemple extrase din corpus. Chiar si în scop autodidactic se poate utiliza un astfel de material, bunăoară, atunci când se caută exemple corecte cu unele cuvinte sau structuri nou învătate, dar încă insuficient exersate. Gilquin și Granger evidențiază această funcție a corpusului (a celui adnotat, în acest caz), în special în ceea ce privește greșelile repetate făcute de către vorbitor, în ciuda interventiilor repetate ale profesorului<sup>385</sup>.

Tot în urma analizei lingvistice a unui corpus al vorbitorilor nonnativi, s-ar putea obține grile de evaluare cu un grad ridicat de validitate, conținând chiar exemple clare preluate din corpus. Or, având în vedere că, la această oră, evaluarea, asemenea predării, se realizează extrem de diferit de la un centru universitar la altul,

<sup>&</sup>lt;sup>383</sup> Gaëtanelle Gilquin, Sylviane Granger, op. cit., 2012, p. 359.

<sup>&</sup>lt;sup>384</sup> Ibidem.

<sup>&</sup>lt;sup>385</sup> Gaëtanelle Gilquin and Sylviane Granger, "How can data-driven learning be use in language teaching?", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, p. 359.

asemenea grile ar putea conduce la standardizarea procesului de evaluare a RLS și la o creștere semnificativă a gradului de obiectivitate. La Departamentul de limbă, cultură și civilizație românească de la Facultatea de Litere a Universității Babeș-Bolyai, se utilizează un set de grile pentru evaluarea producțiilor orale și scrise ale studenților, grile realizate de către membrii departamentului, însă, de multe ori, s-a enunțat necesitatea îmbunătățirii acestora, printr-o nuanțare mai accentuată, bazată, de pildă, pe mai multe exemple de structuri pe fiecare nivel și pe fiecare bandă<sup>386</sup>.

Pe lângă avantajele pe care le poate aduce, din perspectivă didactică, corpusul de producții scrise ale vorbitorilor nonnativi, din perspectiva cercetării RLS, acesta poate constitui, alături de CORLS, un instrument cu o valoare imensă pentru domeniu. El ar putea asigura, prin comparație cu un corpus al vorbitorilor nativi, o analiză mult mai profundă decât cele realizate până la această oră<sup>387</sup>. În plus, subliniem încă o dată că, prin existența și utilizarea acestui material, se va putea valida ceea ce acum doar se crede/se recomandă a fi cunoscut la fiecare nivel<sup>388</sup>. Astfel, o analiză detaliată a erorilor, dar și a structurilor stăpânite de către vorbitorii nonnativi la fiecare nivel poate fi mai simplu efectuată.

De aceea este extrem de relevantă modalitatea în care materialul a fost colectat și introdus în format electronic. "Dacă textele reprezintă producții ale elevilor sau ale copiilor, e important să se păstreze redactarea nonstandard și structurile gramaticale folosite inițial, chiar dacă sunt incorecte, deoarece vor fi extrem de relevante" [tr.n.]<sup>389</sup>. Un alt aspect extrem de important legat de întocmirea corpusului este acela legat de drepturile de autor, precum și aspectele legate de confidențialitate, de exemplu, în cazul producțiilor orale și scrise fiind necesar un acord de confidențialitate, prin care studentul să își exprime acordul sau dezacordul ca producțiile sale să fie folosite în cercetare. La Departamentul de limbă, cultură și civilizație românească, mai sus invocat, există o procedură foarte bine pusă la punct privind confidențialitatea datelor cu caracter personal, care presupune ca, înainte de

<sup>&</sup>lt;sup>386</sup> Despre această temă am vorbit în Dina Vîlcu, Antonela Arieșan, Lavinia Vasiu, "The use of test taker productions in redesigning writing assessment grids. A corpus based study", în *ALTE* (2017). Learning and Assessment: Making the Connections - Proceedings of the ALTE 6th International Conference, 3-5 May 2017, pp. 112-117.

<sup>&</sup>lt;sup>387</sup> Elena Tognini Bonelli, "Theretical overview of the evolution of corpus linguistics", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, p. 25.

<sup>&</sup>lt;sup>388</sup> Lavinia Vasiu, Antonela Arieșan, "RFL – Cohesion. Linking words and phrases (Level B2)", în I. Boldea (ed.), *Globalization and National Identity. Studies on the Strategies of Intercultural Dialogue*, Arhipelag XXI, Târgu Mureș, 2016, pp. 1074-1088.

<sup>&</sup>lt;sup>389</sup> Randi Reppen, "Building a corpus. What are the key considerations?", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, p. 35.

fiecare examen, studenții să completeze un formular cu privire la producțiile elaborate în cadrul acelui examen.

### 4.3. Lingvistica bazată pe corpus. Aplicabilitate interdisciplinară

Pentru a putea avansa, o disciplină are nevoie în permanență de analiză, de reflecție asupra practicilor utilizate și de evidențiere a momentelor când se impune dezvoltarea unor noi resurse, tehnici și direcții. "Lingvistica bazată pe corpus a evoluat extrem de mult pe parcursul ultimilor 20 de ani, extinzându-și și rafinându-și metodele și resursele, pe de-o parte, și lărgindu-și aplicabilitatea interdisciplinară, pe de altă parte" [tr.n.]<sup>390</sup>. Un exemplu în acest sens îl constituie studiul *The online use of violence and journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study*<sup>391</sup>, întreprins de o echipă interdisciplinară formată din lingviști și personal medical, din Lancaster, Marea Britanie. Aceștia au analizat limbajul din mediul online al pacienților suferinzi de cancer și al personalului medical, prin formarea unui corpus alcătuit din metaforele referitoare la violență și la călătorie folosite în mediul virtual de părțile implicate.

De importanță majoră pentru aplicabilitatea interdisciplinară a LC este și analiza efectuată de către Carmen Dyrell și de John Urry, *Mediating climate politics: The surprising case of Brazil*<sup>392</sup>. Autorii au întocmit un corpus nou, Corpusul brazilian pe tema schimbării climatice (*Brazilian Corpus on Climate Change*), care era caracterizat, la momentul apariției, drept un instrument unic în domeniu<sup>393</sup>. Cu ajutorul metodelor LC, autorii au analizat modul în care mass-media din Brazilia a dat naștere unui val de îngrijorare privind schimbarea climatică. Autorii "au arătat cum media a ajutat la remedierea încadrării schimbărilor climatice drept *vreme urâtă*" [tr. n.]<sup>394</sup>. Aceste două studii reprezintă doar o parte din ceea ce LC poate constitui în cercetare.

Acest progres al LC a fost facilitat, pe de-o parte, de avansul tehnologic, iar, pe de altă parte, de preocuparea permanentă a celor din domeniu de a-l dezvolta. În acest fel, apar noi direcții în cercetare. Un prim exemplu îl constituie utilizarea triangulației ca metodă de lucru. Aceasta presupune implicarea mai multor metode

105

<sup>&</sup>lt;sup>390</sup> Dana Gablasova, Vaclav Brezina, Tony McEnery, op. cit., 2019, p. 126.

<sup>&</sup>lt;sup>391</sup> E. Semino et. alli, "The online use of violence and journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study", în *BMJ Supportive & Palliative Care* 7(1), 2017, pp. 60–66. https://doi.org/10.1136/bmjspcare-2014-000785.

<sup>&</sup>lt;sup>392</sup> C. Dayrell, J. Urry, "Mediating climate politics: The surprising case of Brazil", în *European Journal of Social Theory* 18(3), 2015, pp. 257–273. https://doi.org/10.1177/1368431015579962

<sup>&</sup>lt;sup>393</sup> C. Dayrell, J. Urry, *op. cit.*, p. 257.

<sup>&</sup>lt;sup>394</sup> Ibidem.

de analiză a corpusului, posibil și angajarea mai multor lingviști pentru a analiza același corpus. Conform lui Layder, "triangulația metodologică favorizează verificarea validității ipotezelor, ancorează descoperirile în interpretări și explicații mai solide și îi permite cercetătorului să răspundă într-o manieră mai flexibilă unor aspecte neprevăzute sau chiar unor probleme care ar putea apărea în cercetare" [tr. n.]<sup>395</sup>.

O altă direcție este reprezentată de implicațiile pe care lingvistica bazată pe corpus le are în cadrul lingvisticii cognitive sau al psiholingvisticii<sup>396</sup>. Deși, la început, psiholingvistica a fost numită "o disciplină de graniță", "limitrofă" sau un "domeniu interdisciplinar"<sup>397</sup>, în timp, a ajuns să aibă un rol important. Un exemplu în acest sens îl constituie experimentul "asociativ-verbal" (EAV), utilizat în stabilirea vocabularului fundamental necesar predării unei limbi. Astfel, pentru vocabularul fundamental al limbii franceze, s-au utilizat "variante ale acestui experiment, care au constat, de pildă, în a se cere unor elevi de 9-12 ani să scrie primele 20 de cuvinte care le veneau în minte, în asociere cu diverse cuvinte propuse de către profesor"<sup>398</sup>.

Și, nu în ultimul rând, trebuie luată în considerare implicarea statisticii în LC, care, din punctul nostru de vedere, este cea mai importantă și mai utilă direcție adoptată, fără de care nu am putea obține rezultate relevante și nu am putea trage concluzii pentru demararea de noi cercetări. Așa cum arată specialiștii în domeniu, "statistica în LC este despre obținerea unor modele matematice ale unei realități lingvistice complexe. Aceasta ne poate ajuta să descoperim și să elucidăm tipare și tendințe în datele lingvistice, care, în altă situație, ar rămâne nedescoperite" [tr. n.]<sup>399</sup>. O nouă resursă disponibilă sub forma unui corpus nu ar trebui să reprezinte doar o reacție la o analiză imediată de nevoi sau la disponibilitatea unor informații. Aceasta ar putea fi utilizată și ca oportunitate de a contribui la dezvoltarea domeniului, constituind începutul unui proces de implicare în cercetarea problemelor complexe ale acestei discipline. Mai mult, Geoffrey Leach este de părere că, înainte de a

<sup>&</sup>lt;sup>395</sup> D. Layder, *New Strategies in Social Research*, Cambridge, Polity Press, 1993 p.128 apud P. Baker, J. Egbert, (eds.), *Triangulating Methodological Approaches in Corpus-Linguistic Research*, New York, Routledge, 2016, p. 3.

<sup>&</sup>lt;sup>396</sup> Stephan Th. Gries, *Ten lectures on corpus Linguistics with R: applications for usage-based and psycholinguistic research*, Leiden, Boston, Brill, 2020.

<sup>&</sup>lt;sup>397</sup> Tatiana Slama-Cazacu, *Psiholingvistica. O știință a comunicării*, București, Editura All, 1999.

<sup>&</sup>lt;sup>398</sup> *Ibidem*, pp. 691-692.

<sup>&</sup>lt;sup>399</sup> Vaclav Brezina, op. cit., 2018, p. 5.

produce noi materiale, se impune o analiză critică a materialelor existente și propunerea dezvoltării acestora, întrucât fiecare material este perfectibil<sup>400</sup>.

#### 4.4. Concluzii

Materialele destinate limbii române ca limbă străină, chiar dacă sunt declarate ca fiind concepute pentru un anumit nivel din CECRL, sunt elaborate având la bază intuiția vorbitorului nativ, experiența de predare/evaluare în domeniul RLS sau comparația cu materiale elaborate pentru alte limbi. Din cauza acestui demers rezultatele, devenite principalele instrumente disponibile, nu pot garanta unitatea în predare și în evaluare la diferitele centre de limbi din țară și din străinătate. Pentru a se împiedica această situație, se arată că "autorii manualelor și testelor pe nivel să trebui să "uite" că sunt vorbitori nativi și să învețe să configureze microlimba<sup>401</sup> specifică fiecărui palier, chiar dacă, din când în când, vor simți că-și maltratează, cu bună știință, limba" [tr. n.]<sup>402</sup>. Un pas indispensabil înspre atingerea acestui obiectiv îl constituie și un corpus electronic de producții scrise ale vorbitorilor nonnativi, pentru a asigura cercetări riguroase în domeniul limbilor străine.

<sup>&</sup>lt;sup>400</sup> Geoffrey Leech, "New resources, or just better old ones? The Holy Grail of representativeness", în Marianne Hundt, Nadja Husselhauf, Carolin Biewer (eds.), *Corpus Linguistics and the Web*, Amsterdam, Rodopi, 2007, pp. 133-150. https://doi.org/10.1163/9789401203791\_009

<sup>&</sup>lt;sup>401</sup> Vezi *supra*, 4.2.

<sup>&</sup>lt;sup>402</sup> Elena Platon, op. cit., 2019a, p. 73.

# CAPITOLUL V – VOCABULARUL LIMBII ROMÂNE. CADRU GENERAL

### 5.1. Vocabularul limbii române. Considerații generale

Vocabularul sau lexicul unei limbi reprezintă "totalitatea unitătilor lexicale, care există și care au existat, cândva, în limba respectivă"403. Așa cum arată Hristea<sup>404</sup>, precum și Andrei și Ghiță<sup>405</sup>, este unanim acceptat că importanța vocabularului rezultă din faptul că bogăția unei limbi este dată de volumul vocabularului si de varietatea acestuia. Datorită acestui motiv studiul lexicului necesită și merită o atenție sporită. Hristea aduce o serie de argumente în scopul sustinerii acestei afirmatii. Pe de-o parte, acesta arată că schimbările din societate și progresele din stiință și tehnică se reflectă în vocabularul unei limbi, care, consideră autorul, constituie compartimentul cel mai labil si mai deschis influentelor din afară sau chiar împinge și mai departe lucrurile, afirmând că orice vocabular reprezintă, de fapt, o întreagă civilizație. Pe de altă parte, același autor subliniază că "greșelile de ordin lexical sunt mai numeroase și, în general, mai grave decât cele gramaticale"406. O observatie similară face și Stoica, arătând importanta covârsitoare a vocabularului în comunicare. Autoarea susține că "nicio frază corect construită din punct de vedere gramatical și nici un act de vorbire coerent nu se pot realiza fără un bagaj lexical suficient de bogat și de nuanțat însușit" 407. Szudarski arată că există o strânsă legătură între cunoștințele de vocabular pe care un vorbitor le are și competența sa lingvistică în LȚ<sup>408</sup>.

Conform lui Andrei și Ghiță, unitățile lexicale constituie materialul de construcție al unei limbi și sunt reprezentate prin: cuvinte simple, cuvinte compuse, locuțiuni sau chiar și expresii, proverbe și zicători. 409 Iar unitatea de bază a lexicului o constituie *cuvântul*, care e definit drept "asocierea unuia sau a mai multor sensuri cu un complex sau înveliș sonor susceptibil de o întrebuințare gramaticală în

<sup>&</sup>lt;sup>403</sup> Mihail Andrei, Iulian Ghiță, *Limba română. Fonetică, lexicologie, morfosintaxă*, București, Editura Corint, 1996, p. 53.

<sup>&</sup>lt;sup>404</sup> Theodor Hristea, *Introducere în studiul vocabularului*, în Theodor Hristea (coord.), *Sinteze de limba română*, București, Editura Albatros, 1984, p. 7.

<sup>405</sup> Mihail Andrei, Iulian Ghiță, op. cit., 1996, p. 53.

<sup>&</sup>lt;sup>406</sup> Theodor Hristea, op. cit., 1984, p. 8.

<sup>&</sup>lt;sup>407</sup> Gabriela Stoica, op. cit., 2012, p. 182.

<sup>&</sup>lt;sup>408</sup> Paweł Szudarski, op. cit., 2018, p. 34.

<sup>409</sup> Mihail Andrei, Iulian Ghită, op. cit., 1996, p. 54.

procesul comunicării"410. Andrei și Ghită arată că acesta este "o structură fonică la care se asociază un sens, aptă să îndeplinească o funcție în comunicare"411. După Bidu-Vrânceanu si Forăscu, cuvântul "implică aspecte cognitive, un act de cunoaștere, care are loc prin reținerea și redarea proprietăților caracteristice ale obiectului, proces realizat în trepte", Conform GALR, cuvântul are un statut aparte, întrucât acesta "apartine atât sistemului (în calitatea sa de semn lingvistic stabil), cât și comunicării (ca parte a enuntului – text, rezultat al actului enunțiativ)"413. Ca unitate lexicală, ca parte a sistemului, cuvântul aparține vocabularului și "funcționează în virtutea complexelor relații care organizează această parte a limbii"414. Astfel, orice cuvânt se situează în raport cu celelalte prin particularităti privind posibilitatea de asociere, ceea ce îl implică, în egală măsură, atât în organizarea lexicală, cât și în cea gramaticală a limbii<sup>415</sup>. Ca realitate a limbii și ca element care face posibilă comunicarea, cuvântul este văzut drept entitate biplană, asociind "un semnificant (o componentă fonică) cu un semnificat (reprezentat printr-o informație)",416. Se impune a face distincția între cuvânt flexibil (multitudinea de unităti lingvistice = forme care reprezintă o unitate lexicală) si cuvântul-text (care presupune ca fiecare unitate lingvistică să fie un cuvânt-text). În lingvistica modernă, termenul *cuvânt* este înlocuit uneori cu *lexem*<sup>417</sup>, însă acest termen nu este unanim acceptat în comunitatea lingviștilor. La Bidu-Vrânceanu se face diferenta între cuvânt și lexem, în strânsă legătură cu sensul. Astfel, în acceptiunea autoarei, fiecare sens al unui cuvânt polisemantic reprezintă un lexem independent, arătând că "un cuvânt poate fi polisemantic, un lexem nu trebuie să fie"418. Rezultă, așadar, că un cuvânt polisemantic e redat prin atâtea lexeme câte sensuri are.

În cazul achiziției L2, însușirea unui cuvânt nu presupune automat însușirea tuturor sensurilor acestuia, de cele mai multe ori aceasta fiind un proces de durată, care se poate întinde pe parcursul mai multor niveluri. Așadar, putem afirma că,

<sup>&</sup>lt;sup>410</sup> Theodor Hristea, op. cit., 1984, p. 8.

<sup>&</sup>lt;sup>411</sup> Mihail Andrei, Iulian Ghită, op. cit., 1996, p. 54.

<sup>&</sup>lt;sup>412</sup> Angela Bidu-Vrânceanu, Narcisa Forăscu, *Cuvinte și sensuri*, București, Editura Științifică și Enciclopedică, 1988, p. 11.

<sup>&</sup>lt;sup>413</sup> Academia Română, Institutul de Lingvistică "Iorgu Iordan – Al. Rosetti", *Gramatica limbii române. Vol. I. Cuvântul*, București, editura academiei Române, 2005, p. 7.

<sup>&</sup>lt;sup>414</sup> Ibidem.

<sup>&</sup>lt;sup>415</sup> GALR, p. 8.

<sup>&</sup>lt;sup>416</sup> Ibidem.

<sup>&</sup>lt;sup>417</sup> Mihail Andrei, Iulian Ghită, op. cit., 1996, p. 54.

<sup>&</sup>lt;sup>418</sup> Angela Bidu-Vrânceanu, *Structura vocabularului limbii române contemporane. Probleme teoretice și aplicații practice*, București, Editura Științifică și Enciclopedică, 1986, p. 28.

pentru cercetarea noastră, cuvintele polisemantice au un statut aparte, în general, însuşindu-se mai întâi *sensul fundamental*, "adică sensul lexical de bază, principal, esențial al unui cuvânt"<sup>419</sup>, abia apoi, cel(e) *secundar(e)*, "sau adiacent(e), derivat(e) din sensul de bază"<sup>420</sup>. Susținem acest lucru mai ales deoarece, după unii lingviști, aproximativ 80% din volumul lexicului activ al limbii este alcătuit din cuvinte polisemantice<sup>421</sup>, ceea ce face ca o mare parte a lexicului din inventarul nostru să fie inclus în această categorie. Având în vedere accepțiunile prezentate, în analiza noastră vom opera cu termenii vehiculați de Bidu-Vrânceanu, așa cum au fost prezentați mai sus: *cuvânt*, *lexem* și *sens*, precum și *unitate lexicală*, în accepțiunea lui Marello (vezi *supra*, 3.3.).

În general, vocabularul unei limbi nu reprezintă "o masă compactă, omogenă sau nediferențiată" ci e format dintr-un *nucleu al vocabularului* (care conține un număr redus de cuvinte, dar foarte importante) și *masa vocabularului* (care e formată din cuvintele mai rar întrebuințate sau chiar necunoscute de unii vorbitori). De remarcat că, în cazul nucleului vocabularului, de-a lungul timpului s-au folosit mai multe denumiri, precum: *vocabular de bază*, *vocabular fundamental*, *vocabular esențial*, *fond principal lexical*, *fond principal de cuvinte*, *fond lexical uzual* sau chiar *lexic reprezentativ*<sup>423</sup>. Așadar, vocabularul fundamental/fondul principal lexical cuprinde cele mai uzuale cuvinte ale limbii, cele care condiționează într-un fel comunicarea. Acestea sunt, conform lui Hristea, și cele care au cele mai multe derivate sau compuse și care intră în numeroase locuțiuni și expresii.

Așa cum arată Biriș, în ultima perioadă, pe plan internațional, sunt tot mai multe studii importante în care "aspectele lingvistice nu mai sunt predominante, acestea cedând locul chestiunilor de psiholingvistică și aspectelor cognitive ale învățării lexicale" De fapt, perspectiva dinspre științele cognitive demonstrează: "existența unui lexic intern care se construiește prin expunerea la enunțuri orale și scrise de diverse tipuri, lexiconul mental nefiind nicidecum constituit dintr-o listă de

<sup>&</sup>lt;sup>419</sup> Ileana Mureșanu, *Polisemia în limba română*, Cluj-Napoca, casa Cărții de Știință, 2005, p. 22.

 <sup>420</sup> Ibidem.
 421 Vasile Şerban, Ivan Evseev, Vocabularul românesc contemporan, Timișoara, Editura Facla, 1978,

<sup>&</sup>lt;sup>422</sup> Theodor Hristea, *op. cit.*, 1984, p. 13.

<sup>&</sup>lt;sup>423</sup> De remarcat că acest din urmă concept este unul controversat. Unii cercetători consideră că are o sferă mai largă și e opus, pe de-o parte, vocabularului fundamental, iar, pe de altă parte, fondului principal lexical, în vreme ce alți cercetărori nu văd nicio diferență între lexicul reprezentativ și vocabularul de bază.

<sup>&</sup>lt;sup>424</sup> Gabriela Biris, op. cit., 2012, p. 176.

cuvinte, ci dintr-un vast «text» deschis, care presupune interacțiunea complexă a parametrilor psiho-afectivi cu cei contextuali"<sup>425</sup>.

#### 5.2. Perspective în abordarea cunoașterii lexicale în L2

Este cunoscut faptul că o însusire partială a lexicului poate conduce la erori în formularea și transmiterea ideilor, iar "erorile lexicale sunt cele mai frecvente în asimilarea celei de-a doua limbi și sunt considerate cele mai grave, atât de către cei care învată o L2, cât si de către vorbitorii nativi, pentru că ele interferează cu comunicarea"426. Achizitia lexicală are loc în baza cunostintelor deja însusite, iar "un rol deosebit de important îl are profilul lexical pentru limba maternă al vorbitorului în momentul începerii studiului unei noi limbi, deoarece integrarea noilor cunostinte lexicale se face în relație directă cu L1"427. O afirmație similară găsim și la Nation, care arată că există o legătură între limba maternă a unui vorbitor si gradul de efort necesar pentru a-si însusi un cuvânt<sup>428</sup>. Astfel, "povara învătării unui cuvânt este mai grea pentru cei a căror L1 nu se află în relație cu L2" [tr. n.]<sup>429</sup>. Din punctul nostru de vedere, nu doar profilul lexical pentru limba maternă este semnificativ în acest context, ci si acela pentru celelate limbi străine deja cunoscute de către un vorbitor, întrucât toate strategiile și deprinderile dobândite în limbile deja cunoscute vor fi puse în folosință în momentul însușirii lexicului unei noi limbi. Prin urmare, esențială pentru a putea comunica eficient într-o limbă străină se dovedește a fi cunoasterea lexicală. După Schmitt și Schmitt a oferi un răspuns la întrebarea Ce înseamnă a ști un cuvânt? nu presupune o sarcină ușoară<sup>430</sup>, întrucât, așa cum arată Nation, "cuvintele nu sunt unități izolate, ci fac parte din sisteme și niveluri interconectate. Din acest motiv, sunt multe lucruri de stiut despre un cuvânt, în particular, și există diferite niveluri de cunoaștere a acestuia" [tr. n.]<sup>431</sup>.

Conform lui Biriş, "o serie de cercetări din psihologie, psiholingvistică, din științele cognitive în general, s-au concentrat, în ultimii ani, pe chestiunea **învățării** lexicale, a modului în care sunt memorate și stocate cuvintele în lexiconul

. .

<sup>&</sup>lt;sup>425</sup> Mylène Garrigues, *Dictionaire hiérarchiques du français. Principes et méthode d'extraction*, 1992, 93 *apud* Gabriela Biriş, op. cit., 2012, p. 174.

<sup>&</sup>lt;sup>426</sup> P. Meara, *The study of lexis in interlanguage*, 1984 apud Gabriela Stoica, op. cit., 2012, p. 182.

<sup>&</sup>lt;sup>427</sup> Gheorge Doca, "Despre problema dificultăților în procesul învățării limbii române de către străini", în *Predarea limbii române la studenții străini*, 1976, p. 49 *apud* Gabriela Biriș, *op. cit.*, 2012, p. 174.

<sup>428</sup> Este vorba despre *povara învățării (learning burden)*, descrisă de Nation, în *op. cit.*, 2013, pp. 23-24.

<sup>&</sup>lt;sup>429</sup> I. S. P. Nation, op. cit., 2013, p. 24.

<sup>&</sup>lt;sup>430</sup> Norbert Schmitt, Diane Schmitt, *Vocabulary in Language Teaching. Second Edition*, Cambridge University Press, 2020, p. 32.

<sup>&</sup>lt;sup>431</sup> I.S.P. Nation, op. cit., 2013, p. 23.

mental"<sup>432</sup>. Deși accepțiunile au fost inițial destinate teoriei achiziției L1, în timp, și-au dovedit utilitatea și în achiziția L2. Prezentăm, în continuare, trei teorii cunoscute, așa cum apar expuse de către Biriș<sup>433</sup>.

**a. Teoria scenariilor** (*scripts*), a modelelor mentale sau a schemelor, în funcție de autorul la care facem referire, "desemnează acțiuni sau situații de rutină, pe care vorbitorii le cunosc și care au o structură tipică, fiecare schemă antrenând propriul vocabular"<sup>434</sup>. De exemplu, în situația prezentată în figura de mai jos, putem remarca scenariul sau schema care se realizează în mintea vorbitorului în momentul în care întâlnește o anumită situație de comunicare. Cu fiecare etapă din cadrul acestui model mental, se actualizează lexicul necesar bunei îndepliniri a acestei sarcini, dar se poate anticipa și etapa următoare<sup>435</sup>.



Figura 5 - Scenariu mental pentru a lua masa în oraș

**b. Teoria prototipului** "explică modul în care este realizată categorizarea: gruparea entităților din lumea reală se face pe baza unor trăsături dispuse în jurul

112

<sup>&</sup>lt;sup>432</sup> Gabriela Biris, *op. cit.*, 2012, p. 177.

<sup>&</sup>lt;sup>433</sup> Pentru o prezentare mai detaliată a teoriilor, vezi Gabriela Biris, *op. cit.*, 2012, p. 177.

<sup>&</sup>lt;sup>434</sup> Gabriela Biris, op. cit., 2012, p. 177.

<sup>&</sup>lt;sup>435</sup> Ibidem.

unei imagini centrale, prototipice'', plasând la nivelul de bază al categorizării elemente identificabile ușor, din punct de vedere mental, cuvinte simple, frecvente în uz, din perspectivă lingvistică<sup>437</sup>.

c. Conform teoriei învățării lexicale, procesul de achiziție lexicală (achiziționare lexicală, la Biriș<sup>438</sup>) "nu se produce instantaneu, este gradual și relativ lent, pentru că presupune o extindere cantitativă și calitativă a rețelelor memoriei"<sup>439</sup>. Observăm, așadar, în situația acestei teorii intensificarea importanței acordate învățării sensurilor, în detrimental formei cuvintelor. "Estimările cantitative indică faptul că într-o L2 se cunosc în jur de 3000 de cuvinte, după șase ani de studiu, și aproximativ 5000 de cuvinte, după un an de studii academice"<sup>440</sup>.

Dacă aceste teorii caracterizează, în primul rând, achiziția L1, fiind extinsă utilizarea lor și în cazul L2, așa cum s-a arătat mai sus, abordările expuse în continuare sunt elaborate, în principal, pentru achiziția limbilor străine. Prin parcurgerea lor, putem observa complementaritatea care le caracterizează<sup>441</sup>:

1. Distincţia dintre *vocabular receptiv* şi *vocabular productiv* se bazează, conform lui Nation<sup>442</sup>, pe distincţia făcută de Palmer între competenţele receptive şi cele productive<sup>443</sup>. "Termenul *receptiv* implică ideea că receptăm (prin ascultare/citire) un input de la alţi indivizi şi încercăm să îl înţelegem, iar conceptul *productiv* presupune că folosim/producem (prin vorbire/scriere) structuri pentru a transmite mesaje altora" [tr. n.]<sup>444</sup>. În ciuda distincţiei operate între cele două concepte, există, după Nation, şi producere în receptare, întrucât ascultând un mesaj, producem sens<sup>445</sup>. Cu toate că în ultima vreme cele două concepte apar tot mai des în studiile de specialitate, opoziţia *vocabular pasiv* vs *vocabular activ* este, probabil, mai cunoscută şi, în general, este utilizată pentru a desemna acelaşi sens.

<sup>&</sup>lt;sup>436</sup> Gabriela Biris, *op. cit.*, 2012, p. 177.

<sup>&</sup>lt;sup>437</sup> Cuenca și Hilferty, 1999, p. 43 *apud* Gabriela Biriș, *op. cit.*, 2012, p. 177.

<sup>&</sup>lt;sup>438</sup> Gabriela Biris, op. cit., 2012, p. 178.

<sup>&</sup>lt;sup>439</sup> Paul Bogaards, *Le vocabulaire dans l'apprentisage des langues étrangères*, Paris, Didier, 1994, pp. 70-75 *apud* Gabriela Biriş, *op. cit.*, 2012, p. 178.

<sup>&</sup>lt;sup>440</sup> Paul Bogaards, op. cit., 1994, p. 102 apud Gabriela Biris, op. cit., 2012, p. 178.

<sup>&</sup>lt;sup>441</sup> Unele dintre aceste teorii se găsesc prezentate și de către Gabriela Stoica, *op. cit.*, 2012, pp. 182-

<sup>442</sup> I.S.P. Nation, op. cit., 2013, p. 24.

<sup>&</sup>lt;sup>443</sup> Vezi şi Susan M. Gass, Larry Selinker, *Second Language Acquisition. An introductory Course (Third Edition)*, New York/London, Routledge, 2008, pp. 451-452 şi James Milton, *Measuring Second Language Vocabulary Acquisition*, Bristol, Buffalo, Toronto, Multilingual Matters, 2009.

<sup>444</sup> I.S.P. Nation, op. cit., 2013, p. 24.

<sup>&</sup>lt;sup>445</sup> Ibidem.

- **2.** Conform lui Nation, atunci când vorbim despre cunoașterea lexicală, termenii *receptiv* și *productiv* sunt implicați în toate etapele<sup>446</sup>, idee care poate fi observată și în accepțiunea sa privind cunoașterea lexicală, "a ști un cuvânt presupune: formă, sens și utilizare" [tr. n.]<sup>447</sup>. Această abordare, din punctul nostru de vedere, constituie abordarea cea mai de impact. Astfel, după Nation, se disting trei niveluri în cunoașterea unui cuvânt<sup>448</sup>:
- a. (re)*cunoașterea formei*, adică cunoașterea cuvintelor din punct de vedere fonologic și grafic, inclusiv a afixelor;
- b. *cunoașterea sensului* presupune asocierea "unei forme cu un sens și cu un referent/concept"<sup>449</sup>, precum și realizarea de asocieri de cuvinte (ceea ce implică automat faptul că se cunosc contextele de utilizare ale fiecărui element lexical);
- c. cunoașterea restricțiilor de uz reprezintă nivelul la care vorbitorul cunoaște "restricțiile sintactice impuse de termenul în cauză (topică, regim prepozițional etc.) colocațiilor lexicale specifice și restricțiile contextual-semantice de combinare".
- **3.** Pe de altă parte, Ellis<sup>451</sup> susține o separare între învățare *implicită* vs *explicită*, unde "învățarea implicită presupune atenție la stimul, însă nu implică alte operații realizate în mod conștient și e puternic afectată de repetiție. Învățarea explicită este mai conștientă. Vorbitorul nonnativ formulează și testează ipoteze în procesul de căutare a unei structuri, de căutare a regulilor" [tr. n.]<sup>452</sup>.
- **4.** Legat de învățarea implicită, Laufer și Hulstjin<sup>453</sup> au dezvoltat *teoria* gradului de implicare (Involvement Load Hypotesis ILH), "care a avut scopul de a anticipa eficiența activităților de instruire în cazul învățării implicite a vocabularului" [tr. n.]<sup>454</sup> și care implica trei factori: nevoia (need), căutarea (search) și evaluarea și analiza (evaluation). Astfel, se arată că dacă activitățile propuse au un grad ridicat de implicare din partea acestor trei factori, rezultatul învățării va fi mai

<sup>&</sup>lt;sup>446</sup> I.S.P. Nation, *op. cit.*, 2013, p. 26.

<sup>447</sup> Ibidem.

<sup>&</sup>lt;sup>448</sup> Pentru prezentarea acceptiunii lui Nation, vezi I.S.P. Nation, *op. cit.*, 2013, pp. 26-59.

<sup>449</sup> Gabriela Stoica, op. cit., 2012, p. 183.

<sup>450</sup> Ihidem

<sup>&</sup>lt;sup>451</sup> N. Ellis 1995, p. 212 apud I.S.P. Nation, op. cit., 2013, p. 33.

<sup>&</sup>lt;sup>452</sup> N. Ellis 1994, p. 214 apud I.S.P. Nation, op. cit., 2013, p. 34.

<sup>&</sup>lt;sup>453</sup> B. Laufer, J. H. Hulstjin, "Incidental vocabulary acquisition in a second language: The construct of task-induced involvement", în *Applied Linguistics*, 22, 2001, pp. 1-26 *apud* Akifumi Yanagisawa, Stuart Webb, "Involvement Load Hypothesis Plus. Creating an improved predictive model of incidental vocabulary learning", în *Studies in Second Language Acquisition*, 2021, pp. 1-30.

<sup>&</sup>lt;sup>454</sup> Akifumi Yanagisawa, Stuart Webb, *op. cit.*, p. 2.

îmbucurător<sup>455</sup>. Teoria a fost testată de mai mulți cercetători<sup>456</sup>, printre care și Yanagisawa și Webb, care au propus teoria ILH Plus<sup>457</sup>.

**5.a.** O altă abordare e aceea operată de Anderson și Freebody între *breadth* of knowledge (extensia cunoașterii lexicale<sup>458</sup>), "prin care se înțelege numărul cuvintelor pentru care o persoană cunoaște cel puțin o parte dintre cele mai importante sensuri" [tr.n.]<sup>459</sup> și the quality/depth of understanding (profunzimea cunoașterii lexicale<sup>460</sup>), "Presupunem că, în cele mai multe cazuri, are o înțelegere suficient de profundă a unui cuvânt dacă îi oferă toate distincțiile care ar fi în mod normal înțelese de către un vorbitor adult, obișnuit, în circumstanțe normale" [tr. n.]<sup>461</sup>. După Schmitt și Schmitt "size/breath se referă la câte cuvinte sunt cunoscute, în vreme ce depth/quality presupune gradul de cunoaștere a acelor cuvinte" [tr. n.]<sup>462</sup>. Referindu-se la depth of knowledge, Read observă că "pentru cuvintele folositoare și cu frecvență mare, vorbitorii ar trebui să aibă mai mult decât o înțelegere superficială a sensului; ei ar trebui să posede o reprezentare bogată și specifică a sensului, precum și a aspectelor ce țin de sintaxă, colocații, registru etc." [tr. n.]<sup>463</sup>.

**5.b.** Daller et alli vorbesc despre spațiul lexical (*lexical space*), "în care cunoașterea lexicală a unui vorbitor este văzută ca un spațiu tridimensional, unde fiecare dimensiune reprezintă un aspect al cunoașterii unui cuvânt" [tr. n.]<sup>464</sup>. Cele trei axe sunt: *lexical breadth*, *lexical depth* și *fluency*, prin care se înțelege "cât de rapid și automat poate un vorbitor să folosească cuvintele pe care le știe și informațiile pe care le are despre utilizarea acestora"<sup>465</sup>.

După o analiză critică a materialelor destinate predării și învățării RLS, Gabriela Stoica observă că, în cazul acestora, accentul cade asupra extinderii

<sup>&</sup>lt;sup>455</sup> Akifumi Yanagisawa, Stuart Webb, op. cit., p. 2.

<sup>&</sup>lt;sup>456</sup> Aşa cum se arată în Akifumi Yanagisawa, Stuart Webb, *Involvement Load Hypothesis Plus. Creating an improved predictive model of incidental vocabulary learning*, în Studies in Second Language Acquisition, 2021, pp. 1-30.

<sup>&</sup>lt;sup>457</sup> Descrisă în Akifumi Yanagisawa, Stuart Webb, *op. cit.*, 2021, pp. 1-30.

<sup>&</sup>lt;sup>458</sup> Traducerea conceptului îi aparține Gabrielei Stoica, op. cit., 2012, p. 183.

<sup>&</sup>lt;sup>459</sup> Richard C. Anderson, Peter Freebody, "Vocabulary Knowledge", în John T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews*, Newark, Delaware, International Reading Association, 1981, p. 93.

<sup>&</sup>lt;sup>460</sup> Traducerea conceptului îi aparține Gabrielei Stoica, *op. cit.*, 2012, p. 183.

<sup>&</sup>lt;sup>461</sup> Richard C. Anderson, Peter Freebody, op. cit., 1981, p. 93.

<sup>&</sup>lt;sup>462</sup> Norbert Schmitt, Diane Schmitt, op. cit., 2020, p. 38.

<sup>&</sup>lt;sup>463</sup> John Read, "Research in Teaching Vocabulary", în *Annual Review of Applied Linguistics*, 24, 2004, p. 155.

<sup>&</sup>lt;sup>464</sup> Helmut Daller, James Milton, Jeanine Treffers-Daller, "Editor's Introduction: Conventions, terminology and an overview of the book", în Helmut Daller, James Milton, Jeanine Treffers-Daller, *Modelling and Assessing Vocabulary Knowledge*, Cambridge, Cambridge University Press, 2007, pp. 7-8.

<sup>465</sup> Helmut Daller, James Milton, Jeanine Treffers-Daller, op. cit., 2007, p. 8.

cunoștințelor lexicale (*breadh of word knowledge*) în detrimentul profunzimii (*depth of word knowledge*), iar consecința este obținerea unui vocabular pasiv-reactiv, nu a unuia activ-productiv<sup>466</sup>. Astfel, se remarcă nevoia unei structurări mult mai bune a materialului, din perspectivă didactică, pentru a asigura o mai bună însușire a vocabularului.

#### 5.3. Principii de urmat în elaborarea listelor de cuvinte ierarhizate

În producerea inventarelor lexicale, se impun a fi luate o serie de decizii pentru a ne asigura că vom obtine un instrument util si relevant. Conform lui Nation<sup>467</sup>, un aspect important de analizat îl constituie *scopul*, adică este esențial de clarificat dacă inventarul lexical propus este destinat predării sau evaluării, pe de-o parte, iar, pe de altă parte, dacă acesta are utilitate mai mare în producere sau în receptare. Această ultimă perspectivă este una extrem de importantă, întrucât, de exemplu, asa cum evidentiază autorul, dacă vorbim despre receptare, nu e necesar ca fiecare element dintr-o familie lexicală să fie cunoscut, fiindcă sensul unui cuvânt poate fi dedus din context, pe baza cunoștințelor anterioare (forma de bază/radicalul sau sufixele și prefixele care intră în componența sa). Așa cum spunea Anglin, atunci când citim și ascultăm, "putem rezolva probleme de ordin morfologic"<sup>468</sup>. Însă, atunci când folosim cunoștintele lingvistice în producere (scriere sau vorbire), avem nevoie să știm cuvântul, atât forma, cât și înțelesul său, fiindcă pauzele prea lungi sau erorile în utilizarea unor cuvinte pot constitui serioase piedici în comunicare. Dacă e să ne referim la "cuvintele derivate, nu e suficient ca radicalul să fie cunoscut, iar acestuia să i se adauge prefixe și sufixe, după cum credem noi că s-ar putea combina. În acest caz, al cunoștințelor productive, este nevoie ca fiecare element al familiei lexicale să fie cunoscut" [tr. n.]<sup>469</sup>. Şi dacă ne referim, în general, la corpusul folosit pentru a obtine listele ierarhizate pe niveluri, se vor observa inevitabil anumite diferențe între un corpus oral și unul scris. Va exista, pe de-o parte, o gamă mai restrânsă de cuvinte în cazul celui oral, întrucât vorbim despre un tip de discurs ce se construiește pe loc. De asemenea, vor fi și multe repetiții, precum și anumite cuvinte care marchează stilul oral, aspecte ce vor lipsi în cazul corpusului scris. Acestea fiind spuse, considerăm că în cazul cercetării noastre este necesar să avem

<sup>&</sup>lt;sup>466</sup> Gabriela Stoica, *op. cit.*, 2012, p. 183.

<sup>&</sup>lt;sup>467</sup> I.S.P. Nation, op. cit., 2016, p. 8.

<sup>&</sup>lt;sup>468</sup>Anglin, J. M., "Vocabulary development: a morphological analysis", în *Monographs of the Society for Research in Child Development Serial No. 238*, 58 (10 Serial No. 238), 1993, pp. 1–165 *apud* I.S.P. Nation, *op. cit.*, 2016, p. 8.

<sup>&</sup>lt;sup>469</sup> I.S.P. Nation, op. cit., 2016, p. 8.

în vedere atât cunoștințele receptive, cât și pe cele productive, întrucât e important ca studentul să aibă o performantă lingvistică echilibrată.

Potrivit lui Nation, un alt aspect de care trebuie ținut cont și care face ca o listă de cuvinte să fie una valoroasă este *transparența*. Înțelegem prin transparență oferirea de informații despre vocabularul din listă, precum: "dacă este vorba despre un vocabular specific limbii scrise sau celei orale, dacă este cunoscut sau nu de copii etc. În general, transparența e relaționată cu capacitatea listei de a fi reprodusă, însă aceasta poate include și proceduri non-reproductive" [tr. n.]<sup>470</sup>.

Astfel, prima decizie care se impune a fi luată se referă la *ce anume este considerat cuvânt dintr-un corpus* și, în consecință, la ce urmează a fi numărat. De exemplu, combinațiile între litere și numere (de pildă, *A1*, *B2*, *U3*) trebuie să fie numărate drept cuvinte sau este mai indicat să fie izolate și excluse din procesul de numărare? Aceeași întrebare se impune și în cazul unor exclamații care nu au intrat în normă, dar și în situația unor cuvinte care marchează ezitarea, de tipul *hmm*, *aha*, *ăăă* etc. <sup>471</sup>.

O situatie specială o reprezintă substantivele proprii, care ar putea fi numărate separat, dar care presupun cunostințe anterioare<sup>472</sup>. Putem include aici, în primul rând, numele românești de persoane extrem de comune, precum și nume de continente, țări, orașe, poate chiar și unele personalități internaționale (cântăreți, actori, oameni de știință, politicieni, sportivi etc.), care sunt extrem de utile în predare, de exemplu, pentru a oferi exemple ilustrative. Asa cum sustin si Nation si Kobeleva, numele proprii sunt dificil de definit, desi cei mai multi oameni stiu ce reprezintă atunci când le întâlnesc. De obicei, acestea pot fi recunoscute datorită faptului că sunt scrise cu majuscule, însă în analiza noastră trebuie să avem în vedere că nu toate cuvintele scrise cu literă mare sunt substantive proprii. De aceea e important să stabilim ce anume numărăm în cadrul corpusului<sup>473</sup>. Autorii arată că, pe lângă numele de persoane, animale, zei, locuri, evenimente, opere de artă sau nume de firme renumite, apar si unele situatii în care anumite nume de afaceri poate că nu am dori să le numărăm ca substantive proprii, ci să fie contorizate în categoria substantivelor comune, întrucât sunt destul de transparente ca sens. În aceste situații, se impune excluderea lor manuală din lista numelor proprii<sup>474</sup>.

<sup>&</sup>lt;sup>470</sup> I.S.P. Nation, op. cit., 2016, p. 120.

<sup>&</sup>lt;sup>471</sup> *Ibidem*, p. 17.

<sup>&</sup>lt;sup>472</sup> *Ibidem*, p. 18.

<sup>&</sup>lt;sup>473</sup> Paul Nation, Polina Kobeleva, "Proper nouns", în I.S.P. Nation, op. cit., 2016, p. 55.

<sup>&</sup>lt;sup>474</sup> *Ibidem*, p. 56.

În general, nu e nevoie să cunoaștem substantivele proprii înainte de a le întâlni în comunicare, însă Nagy și Andreson<sup>475</sup> au evidențiat o categorie de aproximativ 100 de cuvinte în care sunt necesare cunostinte anterioare, de pildă, în expresii de tipul: Nu e Einstein, nici Gandhi<sup>476</sup>. Datorită acestei observații, apar tot mai des recomandări de a crea separat o listă cu numele proprii care ar urma traseul elementelor lexicale ce trebuie predate și învățate. Aici s-ar include tări, orașe importante, oameni faimosi sau evenimente de renume. De asemenea, este recomandat să existe un nucleu al numelor proprii feminine si masculine<sup>477</sup>. Necesitatea acestui inventar de nume proprii e sustinută si de Brown<sup>478</sup>, care avertizează că nu ar trebui să se presupună că numele proprii nu pot reprezenta o problemă în achizitia limbilor. Si Kobeleva<sup>479</sup> a descoperit că, în cazul ascultării, cel puțin, e esențial ca studenții să aibă cunoștințe anterioare privind numele proprii. Într-o cercetare pe care a întreprins-o, această autoare a observat că au fost îmbunătătiri substanțiale în înțelegerea după auz a noilor texte în două din trei teste de ascultare<sup>480</sup>. Considerăm extrem de pertinentă această observatie, mai ales în cazul începătorilor, întrucât și numele ar reprezenta un element în plus pe care acestia ar trebui să îl deslușească, însă nu este vorba despre un element oarecare, ci, în multe cazuri, unul esential (dacă vorbim, de exemplu, despre numele unei forme de relief sau al unei personalități etc.). În plus, dacă într-un text vor apărea mai multe substantive proprii, pe care studentii nu le vor întelege, receptarea textului va avea foarte mult de suferit, situație ce ar putea fi preîntâmpinată prin întâlnirea cu aceste cuvinte în alte condiții. Din experiența noastră de predare la grupele de studenți străini din programul universitar de An pregătitor, putem spune că unele nume de persoane pot să le pună mari probleme studentilor. Astfel, nume precum Mircea, Mihai, Luca, Horea, Carmen sunt câteva astfel de exemple care pot conduce la confuzii. De multe ori, studenții întâmpină dificultăți în a descoperi, în primul rând, dacă avem de-a face cu un nume, iar, apoi, dacă acesta face referire la un bărbat sau la o femeie. Desigur, această dificultate apare si ca urmare a specificului grupului

<sup>&</sup>lt;sup>475</sup> W. E. Nagy, R. C. Anderson, "How many words are there in printed school English?", în *Reading Research Quarterly*, 19(3), 1984, pp. 304–330 *apud* Paul Nation, Polina Kobeleva, *op. cit.*, 2016, p. 56 <sup>476</sup> Paul Nation, Polina Kobeleva, *op. cit.*, 2016, p. 56.

<sup>&</sup>lt;sup>477</sup> *Ibidem*, pp. 56-57.

<sup>&</sup>lt;sup>478</sup> D. Brown, "An improper assumption? The treatment of proper nouns in text coverage counts", în *Reading in a Foreign Language*, 22, 2010, pp. 355–361 *apud* Paul Nation, Polina Kobeleva, *op. cit.*, 2016, p. 57.

<sup>&</sup>lt;sup>479</sup> P. P. Kobeleva, "Second language listening and unfamiliar proper names: Comprehension barrier?", în *RELC Journal*, 43(1), 2012, pp. 83–98 *apud* Paul Nation, Polina Kobeleva, *op. cit.*, 2016, p. 57. <sup>480</sup> Paul Nation, Polina Kobeleva, *op. cit.*, 2016, p. 57.

nostru țintă, care e compus în mare parte din studenți cu limba maternă arabă, nefamiliarizati cu numele "europene".

În general, putem remarca optiunea fiecărui autor de inventare pentru o metodologie proprie, diferită de a celorlalți. Astfel, unii numără toate elementele lexicale, altii, doar o parte, unii organizează listele după familii lexicale sau câmpuri lexicale. Asadar, e foarte important atunci când facem comparatii să tinem seama care a fost unitatea de măsură. Din punctul nostru de vedere, varianta cea mai accesibilă ar fi să numărăm fiecare element în parte, iar, apoi, dacă e posibil, să se subliste alcătuiască cu anumite grupuri de cuvinte care se analizează/predau/învată/evaluează, de obicei, împreună. Conform lui Nation, acest demers este si cel mai eficient atunci când ne referim la realizarea unui inventar lexical ce are un scop productiv. Pe de altă parte, dacă se elaborează liste de cuvinte în scopuri receptive, nu este atât de important să se numere fiecare cuvânt, arată autorul. În acest ultim caz pot fi inventariate familiile de cuvinte, întrucât se stie că este destul de usor să se obțină un control, cel puțin relativ, asupra înțelegerii acestor cuvinte<sup>481</sup>.

Dang și Webb recomandă ca, în cazul elaborării listelor destinate studenților începători de L2, să nu se folosească familiile de cuvinte, așa cum se procedează de multe ori, cel puțin în cazul limbii engleze, întrucât "cunoștințele morfologice ale acestora pot fi limitate și nu ar fi corect să presupunem că, dacă un membru al unei familii de cuvinte le este cunoscut, ei vor putea recunoaște toate formele sale derivaționale" [tr. n.]<sup>482</sup>. Mai mult, autorii arată că este mai important să li se atragă atenția asupra cuvintelor cu adevărat de frecvență mare în limbă, fiindcă, dacă le vor cunoaște foarte bine pe acestea, mai târziu, când vor avea un bagaj lexical și gramatical suficient de bine dezvoltat, le va fi mult mai ușor să învețe și membrii mai puțin frecvenți ai aceleiași familii de cuvinte. Desigur că, ajungând la acest nivel, vor deține și strategii de decodare a sensului care îi vor ajuta să își însușească cu mai multă ușurință acești noi membri<sup>483</sup>.

În anul 1971, Carroll, Davies și Richman<sup>484</sup>, pentru a realiza dicționarul *Heritage*, au analizat un corpus de cinci milioane de cuvinte obținute din producții școlare, iar, deoarece s-au așteptat să apară probleme legate de ortografie (aspect ce a fost urmărit, de altfel, prin analiza lor, alături de recunoașterea propriu-zisă a

<sup>&</sup>lt;sup>481</sup> I.S.P. Nation, op. cit., 2016, p. 24.

<sup>&</sup>lt;sup>482</sup> Thi Ngoc Yen Dang, Stuart Webb, op. cit., 2016, p. 154.

<sup>&</sup>lt;sup>483</sup> După Thi Ngoc Yen Dang, Stuart Webb, op. cit., 2016, p. 154.

<sup>&</sup>lt;sup>484</sup> J. B. Carroll, P. Davies, B. Richman, *The American Heritage Word Frequency Book*. New York, Houghton Mifflin, Boston American Heritage, 1971.

cuvintelor), aceștia au prevăzut de la început numărarea fiecărei variante a aceluiași cuvânt drept element distinct<sup>485</sup>.

În cazul limbii engleze, de cele mai multe ori se alege familia de cuvinte ca unitate de măsură. Bauer și Nation<sup>486</sup> au realizat un tabel cu principalele afixe, organizate pe niveluri în funcție de importanța lor și de frecvență, cu scopul de a folosi programe informatice pentru a obține automat familiile de cuvinte cele mai frecvente. Ei au propus un model pe șase niveluri, unde primul nivel reprezintă forma de bază, iar, la al doilea nivel, apar o parte dintre cele mai importante flective ce pot fi adăugate acestei forme, urmând ca, la fiecare nivel, să se adauge mai multe afixe (91, în total) astfel încât, până la final, la nivelul șase, să fie formată întreaga familie de cuvinte<sup>487</sup>. Cea mai importantă decizie de luat pare a fi legată de nivelul ce corespunde fiecărui tip de student. Astfel, Dang și Webb susțin că, pentru studenții începători, familiile de cuvinte bazate pe flective sunt cele mai potrivite, întrucât ei nu au cunoștințele gramaticale necesare pentru a înțelege familiile de cuvinte mai extinse care sunt, în general, o combinație între cele mai frecvente cuvinte și cele cu frecvență medie, ele fiind, probabil, irelevante pentru cei aflați la început în învățarea unei limbi<sup>488</sup>.

Dacă, în cazul receptării, de multe ori, sensul cuvintelor poate fi dedus pe baza radicalului sau a afixelor, în ceea ce privește producerea, cea mai bună unitate de măsură pare a fi cuvântul<sup>489</sup>, mai ales că fiecare derivat contribuie la formarea altor colocatii, de aceea este extrem de important ca acestea să fie cunoscute ca atare.

O atenție specială, după Nation, necesită *omonimele*, *omografele*, *omofonele* și cuvintele *polisemantice*. De aceea, în aceste cazuri, e nevoie să se vadă contextul de apariție a acestor cuvinte pentru a evita situația în care, de pildă, este numărat doar un element dintr-o pereche de omonime<sup>490</sup>. Pe de altă parte, în cazul cuvintelor polisemantice, deși există multe voci care sugerează predarea, deci și învățarea, fiecărui sens al unui cuvânt separat, întrucât vorbim de colocații diferite, Nation susține separarea sensurilor unui cuvânt polisemantic în intrări diferite doar în cazul unor sensuri foarte diferite, care presupun automat și învățarea lor separată, fără a se baza pe formă sau pe cunoștințele anterioare<sup>491</sup>.

<sup>&</sup>lt;sup>485</sup> După I.S.P. Nation, op. cit., 2016, p. 17.

<sup>&</sup>lt;sup>486</sup> L. Bauer, I. S. P. Nation, "Word families", în *International Journal of Lexicography*, 6(4), 1993, pp. 253–279.

<sup>&</sup>lt;sup>487</sup> I.S.P. Nation, op. cit., 2016, pp. 26-28.

<sup>&</sup>lt;sup>488</sup> Thi Ngoc Yen Dang, Stuart Webb, op. cit., 2016, p. 154.

<sup>&</sup>lt;sup>489</sup> Vezi J. M. Sinclair, *Looking Up*, London, Collins ELT, 1987 apud I.S.P. Nation, op. cit., 2016, p. 34.

<sup>&</sup>lt;sup>490</sup> I.S.P. Nation, op. cit., 2016, pp. 41-43.

<sup>&</sup>lt;sup>491</sup> *Ibidem*, p. 52.

După Nation, Shin și Grant, o altă decizie care va trebui luată se referă la expresii (multiword units), care pot fi locuțiuni, colocații, sintagme etc., în general, acele expresii care în achiziția L2 se numesc adesea formule memorate, deoarece se predau și se învață ca o unitate (Ce mai faci?/Cu plăcere./Bine ați venit etc.). Acestea sunt extrem de relevante, deoarece reprezintă o parte destul de importantă a bagajului lexical al unui student începător. De aceea, în astfel de situații, autorii arată că, pe lângă o analiză bazată pe calculator, se impune și judecata din partea factorului uman. Atunci când se face analiza de corpus, este foarte important ca numărarea elementelor din aceste expresii să se facă doar o singură dată, sau ca expresii, sau ca elemente lexicale individuale, altfel valorile obținute vor fi irelevante<sup>492</sup>. Totuși, la nivelul A1, când cele mai multe elemente sunt memorate, considerăm că e dificil să facem această separare, mai ales că multe cuvinte care formează expresiile sunt utilizate și individual.

Si *acronimele*, dar și *abrevierile*, spune Nation, ar putea fi incluse într-un inventar separat, întrucât au o situație specială. Vorbim aici despre cuvinte, precum: *SUA*, *UK*, *NASA*, *ex.*, *etc.*, *kg*, *km*, *m*, *cm* ș.*a*.<sup>493</sup>. Considerăm că și această propunere este una utilă și aplicabilă și în cazul analizei noastre.

O situație specială o au *cuvintele de legătură* (așa-numitele *function words*), prepozițiile, conjuncțiile, articolele, verbele auxiliare etc. Și în acest caz, Nation susține că ar trebui constituită o sublistă cu aceste cuvinte, având în vedere că multe dintre ele trebuie învățate într-un alt mod decât cuvintele cu conținut lexical, bazându-ne mai mult pe utilizare și pe cunoștințe gramaticale. În plus, acestea sunt și extrem de frecvente<sup>494</sup>. Noi optăm totuși pentru păstrarea lor în lista principală, cel puțin pentru acest nivel.

#### 5.4. Criterii

Analizând materialele disponibile legate de tema noastră, am putut observa că, deși uneori frecvența pare a fi singurul criteriu care determină elementele unei liste de cuvinte, acest lucru nu este întru totul adevărat. Astfel, în funcție de situație, frecvenței i se alătură *utilitatea*, *răspândirea* (*range*) și *dispersia*, criterii pe care le vom discuta mai jos. Pe lângă listele create, în primul rând, pornind de la criteriul frecvenței, mai există o categorie, pe care doar o amintim aici, realizată *pe baza* 

<sup>&</sup>lt;sup>492</sup> Vezi Paul Nation, Dongkwang Shin, Lynn Grant, "Multiword units", în I.S.P.Nation, *op. cit.*, 2016, pp. 71-79.

<sup>&</sup>lt;sup>493</sup> I.S.P. Nation, op. cit., 2016, pp. 85-87.

<sup>&</sup>lt;sup>494</sup> *Ibidem*, p. 89-91.

capacității cuvintelor de a acoperi ideile exprimate într-o limbă, precum și a celei de a se combina în așa fel încât cu ajutorul lor să se poată exprima tot ce este nevoie. În engleză este vorba despre Basic English, ce conține doar 850 de familii de cuvinte<sup>495</sup>. După cum afirmă și Nation, majoritatea acestor cuvinte sunt cuvinte foarte frecvente, însă nu toate cuvintele frecvente sunt incluse în această listă<sup>496</sup>.

Prezentăm, în continuare, cele trei criterii *statistice* folosite, de obicei, în elaborarea listelor de cuvinte:

#### 5.4.1. Răspândirea

Așa cum afirmă și Nation<sup>497</sup>, dacă am fi nevoiți să stabilim un clasament al criteriilor, acesta ar fi cel mai important criteriu, deoarece răspândirea arată în câte texte diferite apare un anumit cuvânt. Utilitatea acestui criteriu poate fi observată în situațiile în care, într-un text, un anumit cuvânt este extrem de des întâlnit, în vreme ce, în celelalte, acesta nu apare. Totuși, este de avut în vedere că acest criteriu e influențat de dimensiunile (sub)corpusului. "Cu cât un (sub)corpus e mai mare, cu atât un cuvânt de frecvență redusă are mai mari șanse să apară. De aceea, pentru a avea rezultate valide, e necesar ca fiecare (sub)corpus să aibă aceeași dimensiune și să fie coerent, adică să conțină texte similare ca tip" [tr. n.]<sup>498</sup>. Ceea ce e demn de reținut este faptul că răspândirea și frecvența sunt relaționate. Astfel,

"dacă avem un corpus compus din mai multe subcorpusuri mari, iar un cuvânt va apărea în toate aceste subcorpusuri, însă, în unele dintre ele, doar o dată sau de două ori, vom presupune automat un grad de răspândire mare, ceea ce ne-ar face să afirmăm că avem de-a face cu un cuvânt cu frecvență ridicată, deși nu este cazul. În această situație, e important să recurgem și la cel de-al treilea criteriu, cel al dispersiei, pentru a evita o clasificare eronată" [tr. n.]<sup>499</sup>.

#### 5.4.2. Frecvența

Aceasta ilustrează cel mai bine importanța pe care o are fiecare cuvânt în comunicare. E esențial ca elementele lexicale cu frecvență mare în limbă să fie printre primele cuvinte învățate, după Zimmerman, întrucât "acestea vor apărea, în

<sup>&</sup>lt;sup>495</sup> C. K. Ogden, *The Basic Words*, London, Kegan Paul, Trench, Trubner & Co., 1932, disponibil online aici: http://ogden.basic-english.org/texts/tbw.html.

<sup>&</sup>lt;sup>496</sup> I.S.P. Nation, op. cit., 2016, p. 13.

<sup>&</sup>lt;sup>497</sup> *Ibidem*, p. 120.

<sup>&</sup>lt;sup>498</sup> Ibidem.

<sup>&</sup>lt;sup>499</sup> Ibidem.

mod natural, atât în cadrul orelor, cât și în afara acestora, oferindu-le vorbitorilor nonnativi expunerea repetată la un cuvânt, de care aceștia au nevoie" [tr. n.]<sup>500</sup>. Mai mult, aceeași autoare susține că doar unele dintre cuvintele unei limbi "merită să li se acorde timp prețios în cadrul orei" [tr. n.]<sup>501</sup>. Este vorba despre cuvintele cu frecvență mare "care sunt legate de teme utile și care le pot pune probleme deosebite" [tr. n.]<sup>502</sup> vorbitorilor nonnativi.

Prin frecvență, se "încearcă o stratificare a cuvintelor de importanță diferită în vocabular"<sup>503</sup>. Frecvența este cea care "indică de câte ori se întrebuințează un cuvânt într-un text dat; cu cât un cuvânt apare mai mult, cu atât frecvența e mai mare, și invers"<sup>504</sup>. Tom Cobb explică motivele pentru care ar trebui să se învețe, în limbile străine, mai întâi, cele mai frecvente cuvinte: "De ce ai dori să înveți cuvintele frecvente din engleză? Din simplul motiv că engleza, asemenea oricărei alte limbi, are obiceiul de a *recicla*, la nesfârșit, un număr relativ mic de cuvinte, iar, dacă știi aceste cuvinte, atunci capacitatea ta de citire se poate dezvolta impresionant cu un efort relativ mic." [tr. n.]<sup>505</sup>. Unele cuvinte strâns legate de tema unui text pot fi foarte frecvente, situație ce ar putea conduce spre un rezultat neconcludent. De aceea e important să relaționăm criteriile și, astfel, cel mai probabil se va observa că acel cuvânt are o răspândire redusă, ceea ce îl va exclude din lista cuvintelor foarte frecvente<sup>506</sup>.

#### 5.4.3. Dispersia

"Dispersia măsoară cât de uniform sunt răspândite ocurențele unui cuvânt în mai multe texte" [tr. n.]<sup>507</sup>. Acest criteriu combină frecvența și răspândirea și verifică dacă ocurențele unui cuvânt în diferite subcorpusuri sunt echilibrate. Nation arată care este diferența între răspândire și frecvență, pe de-o parte, și dispersie, pe de altă parte. Primele se măsoară prin numărare, arătând în câte subcorpusuri apare un cuvânt, respectiv, cât de des apare un cuvânt, în vreme ce dispersia e calculată cu

<sup>&</sup>lt;sup>500</sup> Cheryl Boyd Zimmerman, op. cit., 2009, p. 7.

<sup>&</sup>lt;sup>501</sup> *Ibidem*, p. 7.

 $<sup>^{502}</sup>$  Ibidem.

<sup>&</sup>lt;sup>503</sup> Bidu-Vrănceanu, Angela et alli, op. cit., 2001, p. 224.

<sup>&</sup>lt;sup>504</sup> Ibidem

<sup>&</sup>lt;sup>505</sup> Tom Cobbs, *Why & how to use frequency lists to learn words*, <a href="http://www.lextutor.ca/research/">http://www.lextutor.ca/research/</a>, ultima consultare, în 20.11.2022.

<sup>&</sup>lt;sup>506</sup> După I.S.P. Nation, op. cit., 2016, p. 120.

<sup>&</sup>lt;sup>507</sup> *Ibidem*, p. 6.

ajutorul unei formule care ține cont și de frecvență, și de răspândire, "adică frecvența unui cuvânt în fiecare subcorpus" [tr. n.]<sup>508</sup>.

Întrucât cele trei criterii sunt asemănătoare, pot fi folosite toate sau se poate renunța la unul dintre ele<sup>509</sup>. După oricare dintre cele două opțiuni (folosirea a două sau a trei criterii), se obțin două sau trei valori care trebuie să fie ierarhizate. "Pentru a obține rezultate consistente, specialiștii le-au combinat într-o formulă cu ajutorul căreia se obține un număr ce e folosit în clasificarea cuvintelor" [tr. n.]<sup>510</sup>. După Nation, nu întotdeauna se folosesc formule, însă, atunci când este nevoie să se ia niște decizii, în funcție de scopul pentru care se elaborează lista, formulele pot fi utilizate<sup>511</sup>. Astfel, dacă avem nevoie de o listă care să arate utilitatea generală, de exemplu, se va ține cont în primul rând de criteriul răspândirii și al frecvenței<sup>512</sup>.

Conform lui Nation<sup>513</sup>, există două direcții atunci când se elaborează liste de cuvinte bazate pe corpus. Prima este de a folosi cele trei criterii prezentate mai sus, *răspândire, frecvență și dispersie,* iar cea de-a doua opțiune presupune să se folosească o metodă statistică similară primei direcții, însă să existe și posibilitatea de a ajusta rezultatele, utilizând și alte criterii, astfel încât să ne asigurăm că anumite grupuri de cuvinte, care sunt predate și învățate, de obicei, împreună, vor fi prezente și în aceste liste tot împreună. De exemplu, așa-numitele *lexical sets*, care constituie categorii de cuvinte, precum: numerele, zilele săptămânii, lunile anului, gradele de rudenie sau un vocabular de supraviețuire<sup>514</sup>.

Fiecare direcție are avantajele sale, așa cum arată Nation<sup>515</sup>. Dacă se optează pentru prima variantă, ne asigurăm că, într-o altă situație, folosind același corpus, se va obține exact același rezultat. Cea de-a doua direcție, introducând factorul subiectiv în luarea deciziei, oferă posibilitatea utilizatorului de a obține o listă mult mai utilă și mai completă, mai aproape de ce are acesta nevoie, însă nu poate fi ușor replicată de altcineva, ceea ce nu poate fi considerat neapărat un punct slab, întrucât cel mai important este să se obțină un material cât mai complet. De reținut este faptul că:

<sup>&</sup>lt;sup>508</sup> I.S.P. Nation, *op. cit.*, 2016, p. 121.

<sup>&</sup>lt;sup>509</sup> Situație mai frecvent întâlnită.

<sup>&</sup>lt;sup>510</sup> I.S.P. Nation, op. cit., 2016, p. 121.

<sup>&</sup>lt;sup>511</sup> Formulele utilizate în statistica lingvistică sunt prezentate în capitolul 6.

<sup>&</sup>lt;sup>512</sup> I.S.P. Nation, op. cit., 2016, p. 122.

<sup>&</sup>lt;sup>513</sup> *Ibidem*, pp. 118-119.

<sup>&</sup>lt;sup>514</sup> P. Nation, & D. Crabbe, "A survival language learning syllabus for foreign travel", în *System*, 19(3), 1991, pp. 191–201.

<sup>&</sup>lt;sup>515</sup> I.S.P. Nation, op. cit., 2016, p. 119.

"o listă de cuvinte nu reprezintă produsul final, ci constituie un instrument ce va putea fi folosit pentru elaborare de curriculum, predare și testare. Calitatea și relevanța listei de cuvinte sunt cele mai importante calități, iar acestea pot fi obținute doar cu implicarea unor proceduri ce nu pot fi ușor reproduse. Aceste proceduri includ: completarea grupurilor de cuvinte, atenție sporită acordată nevoilor grupului-țintă și corectarea eventualelor deficiențe în corpus" [tr.n.]<sup>516</sup>.

Semnificativ este și faptul că unele dintre cele mai utile liste de cuvinte pentru limba engleză au fost realizate urmând cea de-a doua direcție<sup>517</sup>. Din punctul nostru de vedere, cea de-a doua perspectivă este mult mai complexă, motiv pentru care vom opta pentru aceasta în elaborarea inventarului nostru.

Pe lângă criteriile prezentate deja, mai există o serie, din punctul nostru de vedere extrem de utile, care ar trebui folosite în scopul obținerii unui inventar lexical cât mai relevant:

## 5.4.4. Criteriul întregirii sferelor lexicale

După Nation<sup>518</sup>, o modalitate de a adapta o listă și de a o face cât mai eficientă este să ne asigurăm că ea include toți membrii unui sfere lexicale (*closed lexical set*). "Un grup închis de cuvinte este constituit din acele elemente lexicale care se potrivesc în general împreună sub un cuvânt-titlu și se completează destul de ușor, cu un număr relativ mic de cuvinte" [tr. n.]<sup>519</sup>. Astfel, vorbim despre grupuri compuse din: numere, zilele săptămânii, lunile anului, anotimpuri, puncte cardinale, relații de rudenie, culori, saluturi și cuvintele de politețe. Desigur că aceste grupuri pot fi completate cu altele, specifice vârstei grupului-țintă, așa cum subliniază Nation. În cazul în care lucrăm cu copii, de pildă, vom include, probabil, animale, jocuri și jucării<sup>520</sup>.

Într-un astfel de demers ar putea apărea o întrebare, de altfel, cât se poate de pertinentă: de ce am avea nevoie de această intervenție dacă ne bazăm doar pe analiza empirică a corpusului? Or, tocmai fiindcă ne bazăm pe corpus este posibil ca unele elemente din aceste sfere lexicale să nu apară în lista noastră. Și nu pentru că acestea nu ar fi frecvente sau utile, ci pentru că un student care a redactat un text sau a participat la o conversație nu ar fi putut nicidecum să folosească toate elementele,

<sup>&</sup>lt;sup>516</sup> I.S.P. Nation, op. cit., 2016, p. 119.

<sup>&</sup>lt;sup>517</sup> Este vorba, de exemplu, despre *General Service Lists*, *Basic English* și *Survival Vocabulary for Foreign Travel*.

<sup>&</sup>lt;sup>518</sup> I.S.P. Nation, *op. cit.*, 2016, pp. 122-123.

<sup>&</sup>lt;sup>519</sup> *Ibidem*, p. 122.

<sup>&</sup>lt;sup>520</sup> *Ibidem*, p. 123.

mai ales că unele se exclud reciproc sau necesită selecție în funcție de situația de comunicare.

Un alt argument vine tot de la Paul Nation, care a observat în analizele sale că, de exemplu, zilele săptămânii se aflau, majoritatea, în primele 2000 de cuvinte, cu excepția zilelor *marți* și *joi*. Aceeași situație a întâlnit-o și în cazul lunilor anului, toate lunile anului, fără *februarie* și *august*, regăsindu-se în primele cele mai frecvente 2000 de cuvinte<sup>521</sup>. Așadar, situațiile de acest tip ne arată că e important să întregim listele cu aceste elemente pentru a putea avea un tablou cât mai complet. Ar fi ineficient și ilogic să separăm pe două niveluri lunile anului, de pildă. De fapt, ar fi de-a dreptul imposibil, întrucât oricine învață limba română are nevoie să știe fiecare lună, deoarece le va întâlni pe toate, cu siguranță, în varii contexte, ca, de exemplu, în discuții despre dată, zile de naștere, anotimpuri etc. Și din perspectiva celui care învață o L2/LS, așa cum arată Zimmerman, "e mai dificil să învețe lunile anului într-o listă decât să le învețe una după cealaltă, făcând conexiuni cu informații familiare" [tr. n.]<sup>522</sup>, precum: activitățile făcute, în general, persoane născute în acea lună, evenimente importante din viața noastră care s-au petrecut în luna respectivă etc.

"Sferele lexicale (*lexical sets*) par a fi mai mult asociate cu anumite tipuri de texte. În studiul lui Sorell s-a observat că zilele săptămânii au apărut în primele 500 de cuvinte folosite în conversații, iar lunile apăreau în primele 1000 de cuvinte necesare pentru scriere generală, în vreme ce culorile erau mai bine reprezentate în textele narative și descriptive" [tr. n.]<sup>523</sup>. Există totuși specialiști care nu recomandă utilizarea acestui criteriu. Printre ei se numără chiar Nation, în unele situații, dar și West. Ambii susțin că, prin întregirea acestor grupuri de cuvinte, nu facem decât să includem cuvinte cu frecvență mică printre cuvintele cu frecvență ridicată, ceea ce va face ca altele mai frecvente să nu fie incluse sau să fie incluse pe poziții inferioare, fapt ce aduce cu sine și necesitatea de a preda și deci de a învăța cuvinte, care nu sunt la fel de importante<sup>524</sup>. Cu toate că acest argument este, într-adevăr, demn de luat în seamă, amintim că frecvența nu este singurul criteriu folosit în elaborarea inventarelor și că, pentru a obține instrumente relevante, unele cuvinte sunt important de reținut, în ciuda rangului pe care îl dețin în lista de frecvență.

21 - -

<sup>&</sup>lt;sup>521</sup> I.S.P. Nation, op. cit., 2016, p. 123.

<sup>&</sup>lt;sup>522</sup> Cheryl Boyd Zimmerman, op. cit., 2009, p. 19.

<sup>&</sup>lt;sup>523</sup> I.S.P. Nation, op. cit., 2016, p. 123.

<sup>&</sup>lt;sup>524</sup> *Ibidem*, p. 124.

#### 5.4.5. Eficiența

Un alt criteriu ce ar putea fi folosit este criteriul eficienței, folosit de către West<sup>525</sup>, în 1953, atunci când a elaborat *General Service List*. Astfel, așa cum arată Nation<sup>526</sup>, pentru a obține o listă eficientă, pe lângă rezultatele statistice, West a utilizat alte subcriterii: *cel al ușurinței sau al dificultății de învățare*, subcriteriul *necesității, al suprapunerii cu cuvinte deja în listă*, dar și *cel al stilului* sau *al cuvintelor care exprimă intensitatea și emoția*. Criteriul ușurinței sau al dificultății de învățare a impus includerea în listă a cuvintelor mult mai ușor de învățat, deoarece erau cuvinte internaționale, împrumuturi sau cel puțin conțineau sau puteau fi relaționate cu unele cuvinte deja cunoscute. "Unele cuvinte pot să nu fie foarte frecvente, însă reprezintă singura modalitate de a exprima ceva important"<sup>527</sup>.

În cadrul unor studii anterioare tezei noastre<sup>528</sup>, am propus și noi o serie de criterii pentru selecția elementelor care ar urma să facă parte dintr-un inventar lexical pe nivelurile comune de referință din CECRL. Printre acestea se numără și criteriul utilității, care se referă la lexicul minimal pe care considerăm că are nevoie să și-l însușească un individ ce învață o limbă străină pentru a putea lua parte activ la o situație de comunicare, precum și la conceptele, termenii relaționali, cuvintele care compun cerințele activităților sau cele necesare formulării unor întrebări.

Astfel, în momentul stabilirii inventarelor lexicale, după aplicarea celorlalte criterii, va fi, cu siguranță, necesar să decidem dacă nu ar trebui incluși aici și alți termeni (care, din alte puncte de vedere, nu ar aparține nivelului), din rațiuni de stringentă necesitate. Un exemplu în acest sens ar putea fi reprezentat de unele forme verbale, necesare pentru înțelegerea cerințelor: *citește!*, *rezolvă! potrivește!* sau de cuvinte, precum: *adevărat/fals, perechi, grupuri* etc. ). Gairns și Redman, în *Working with words. A guide to teaching and learning vocabulary*, susțin că sub acest criteriu

<sup>&</sup>lt;sup>525</sup> Un aspect extrem de interesant, demn de reținut, este folosirea unor criterii pentru a exclude anumite cuvinte din listă. Este cazul criteriului *stilului și al cuvintelor care exprimă intensitatea și emoția*, folosit tot de West în elaborarea listei sale generale de cuvinte, pentru a evita inventarierea cuvintelor asociate cu stilul specific literaturii.

<sup>&</sup>lt;sup>526</sup> I.S.P. Nation, op. cit., 2016, p. 125.

<sup>527</sup> Ibidem.

<sup>528</sup> Antonela-Carmen Suciu, "Strategii de predare a competenței lingvistice (lexic, fonologie, morfologie și sintaxă) – Gramatică și vocabular", în Elena Platon, Diana-Viorela Burlacu, Ioana Sonea (coord.), *Procesul de predare-învățare a limbii române ca limbă nematernă (RLNM) la ciclul primar. RLNM: P1-ciclul primar*, Cluj-Napoca, Editura Casa Cărții de Știință, 2011, p. 83. și Antonela Arieșan, "Inventarul lexical al limbii române ca limbă străină – Propunere de cercetare", în Elena Platon, Antonela Arieșan (eds.), *Actele conferinței aniversare "40 de ani de limba română ca limbă străină la UBB. 1974-2014*", Cluj-Napoca, Editura Casa Cărții de Știință, 2014, pp. 248-250.

<sup>&</sup>lt;sup>529</sup> Conform *Descrierii minimale a limbii române ca limbă străină*. *A1*, *A2*, *B1*, *B2*, verbele la modul imperativ se învată la nivelul A2, însă în cerințele exercitiilor apar de la A1.

ar trebui să se includă și terminologia gramaticală, care l-ar ajuta pe cel care învață limba să folosească cu o mai mare ușurință și dicționarul sau alte materiale<sup>530</sup>. Şi Zimmerman arată că pe lângă cuvintele frecvente, în general, e important să avem în vedere și cuvintele care sunt frecvente în cadrul orelor noastre<sup>531</sup>.

Considerăm în continuare util acest criteriu și ne propunem alcătuirea unui microcorpus format din cerințe/sarcini de lucru pentru nivelurile A1 și A2, preluate din materialele de predare/evaluare, în scopul obținerii unei liste ce va fi mai apoi încrucișată cu inventarul lexical bazat pe analiza corpusului de producții scrise ale nonnativilor. Considerăm ca această listă ar putea constitui și singură un instrument util atât pentru cercetători și cadre didactice, cât și pentru un student înscris într-un program de studiu sau pentru un autodidact.

#### 5.4.6. Criterial subjectiv

Conform lui Nation, pe lângă aceste criterii, se impune să avem în vedere caracteristicile utilizatorului, întrucât, oricât de bine am alcătui un corpus, nu vom reuși să obținem un material care să reprezinte în totalitate nevoile grupului nostru țintă. De aceea "este posibil să fie necesar să adăugăm anumite cuvinte, despre care știm foarte clar că sunt relevante pentru utilizator, dar și să excludem altele" [tr. n.]<sup>532</sup>. Nation arată că este extrem de dificil să dispunem de un corpus suficient de mare și de reprezentativ pentru tot ce ne dorim. De aceea propune folosirea unui corpus mai puțin extins, dar relevant pentru cercetarea desfășurată aici. Desigur că, în acest caz, vom fi nevoiți să implicăm și criterii subiective în analiza noastră. Justificarea pentru introducerea acestor criterii subiective ar sta în faptul că un corpus mai mic se bucură de mai puțină încredere, însă "atâta timp cât ajustările sunt minore, iar numărul lor se încadrează în valoarea obișnuită a variației în acel caz și acestea sunt evidențiate în listă, ajustările ar trebui să fie acceptabile" [tr. n.]<sup>533</sup>. Pe lângă aceste măsuri, Nation recomandă, acolo unde este posibil, compararea listelor cu alte liste similare, în cazul nostru, din alte limbi<sup>534</sup>.

<sup>&</sup>lt;sup>530</sup> Ruth Gairns și Stuart Redman, *Working with words. A guide to teaching and learning vocabulary*, University of Cambridge, Cambridge, 1986, pp. 61-62.

<sup>&</sup>lt;sup>531</sup> Cf. Cheryl Boyd Zimmerman, Word Knowledge. A Vocabulary Teacher's Handbook, Oxford University Press, 2009, p. 7.

<sup>&</sup>lt;sup>532</sup> I.S.P. Nation, *op. cit.*, 2016, p. 125.

<sup>&</sup>lt;sup>533</sup> *Ibidem*, p. 126.

<sup>&</sup>lt;sup>534</sup> Ibidem.

#### 5.4.7. Capacitatea combinatorie

Acest criteriu face parte din setul de criterii propuse de noi<sup>535</sup>. Capacitatea combinatorie a cuvintelor reprezintă o "aptitudine a oricărei unități lingvistice de a stabili relații de succesiune (sintagmatice) cu altă unitate (alte unități) în vederea obținerii unor combinații mai largi"<sup>536</sup>. Astfel, considerăm util acest criteriu, întrucât, cu ajutorul său, ne putem asigura de includerea acelor termeni care intră în relație cu un număr important de cuvinte (de pildă, verbe ca *a avea, a fi, a face*, dar și substantive ori adjective relevante). Acest criteriu va facilita formarea a numeroase enunțuri cu un număr relativ redus de cuvinte, având o utilitate mare, mai ales la nivelurile inferioare, A1 și A2. De observat că acest criteriu intră sub incidența factorului subiectiv, lucru pe care ni-l asumăm.

# 5.5. Delimitarea numărului de cuvinte pe fiecare nivel de competență lingvistică

Există o categorie de cuvinte care nu apar frecvent în comunicare, astfel că stabilirea unei granite între cuvintele cu frecvență ridicată și cele cu frecvență redusă se va face într-un mod arbitrar<sup>537</sup>. Dacă ne referim din nou la cazul limbii engleze, aici apar două clasificări bazate pe frecvență. Pe de-o parte, una se găsește la Nation, care face deosebirea între cuvintele cu frecvență ridicată (primele 2000 de cuvinte, cu familiile lor de cuvinte) și cele cu frecvență redusă (celelalte), iar, pe de altă parte, Schmitt si Schmitt<sup>538</sup> propun, în special din ratiuni didactice, o clasificare bazată pe trei categorii: frecvență ridicată (primele 3000 de cuvinte, cu familiile lor de cuvinte), frecvență moderată/medie (6000 de cuvinte, cu familiile de cuvinte) și frecventă redusă (restul)<sup>539</sup>. În ceea ce privește stabilirea limitei (cut-off point) între cuvintele cele mai frecvente si cele mai putin frecvente, observăm că decizia nu este ușor de luat, întrucât aceasta depinde de cele mai multe ori de corpusul avut la dispoziție, care, oricât de bine ar fi fost alcătuit, tot nu poate îndeplini toate cerințele, mai ales atunci când vorbim despre obținerea unui material care va avea mai multe utilizări, cum e cazul materialului propus de noi<sup>540</sup>. De aceea, sustine Nation, e important să intervină și factorul subiectiv în elaborarea listelor, un ochi critic care poate remarca lipsa unor cuvinte sau necesitatea completării unor seturi de cuvinte

<sup>&</sup>lt;sup>535</sup> Antonela-Carmen Suciu, *op. cit.*, 2011, p. 83, și Antonela Arieșan, *op. cit.*, 2014, pp. 248-250.

<sup>&</sup>lt;sup>536</sup> Angela Bidu-Vrănceanu et alli, *op. cit.*, 2001, p. 115.

<sup>&</sup>lt;sup>537</sup> După I.S.P. Nation, op. cit., 2016, p. 4.

<sup>&</sup>lt;sup>538</sup> Schmitt, N., & Schmitt, D., "A reassessment of frequency and vocabulary size in L2 vocabulary Teaching", în *Language Teaching*, 47(4), 2014, pp. 484–503, *apud* I.S.P. Nation, *op. cit.*, 2016, p. 4. <sup>539</sup> După I.S.P. Nation, *op. cit.*, 2016, p. 4.

<sup>&</sup>lt;sup>540</sup> *Ibidem*, pp. 19-22.

(zilele săptămânii, numere, anotimpuri etc.). Acest ochi critic poate fi reprezentat chiar de către autorul listei, care, pe baza experienței, poate ajusta lista, sau un alt expert sau grup de experți, care validează și/sau completează inventarul. Pentru situația în care se elaborează o listă pentru predare/învățare, atunci valoarea frecvenței ar trebui să fie mai mare de unu<sup>541</sup>.

În cazul cercetării noastre, pentru a putea obține listele de cuvinte pentru fiecare nivel de competență lingvistică, este nevoie să stabilim limita dintre niveluri. Desigur că rezultatele analizei corpusului vor fi, în primul rând, cele care vor dicta această limită, însă considerăm că e important să ne stabilim și unele așteptări privind dimensiunea inventarului pe fiecare nivel avut în vedere. Astfel, o primă delimitare se poate face ținând cont de descriptorii din CECRL privind volumul și stăpânirea vocabularului, însă aceste descrieri sunt mult prea generale pentru a ne fi de folos în stabilirea unei *granițe* între niveluri. Oricum, chiar și atunci când această limită se poate fixa, ea este atât de fină încât poate fi cu ușurință depășită, în funcție de situația de învățare<sup>542</sup>, dar devine extrem de utilă, atunci când vine vorba despre evaluare sau autoevaluare.

Din studiile consultate, reiese că numărul de cuvinte pentru fiecare nivel poate fi calculat în două moduri: pe de-o parte, plecând de la situația celorlalte limbi, iar, pe de altă parte, folosind informațiile legate de numărul de cuvinte care pot fi învățate într-o zi, precum și pe cele despre numărul de ore aferente unui nivel.

**a.** În continuare vom prezenta o statistică privind numărul de cuvinte pe fiecare nivel, preluată din articolul *Europe and language learning: The challenges of comparable assessment*, de Françoise Kusseling și Wilfried Decoo<sup>543</sup>. Am adăugat acestui tabel, în ultimele două rânduri, informații despre limbile franceză și greacă:

Tabelul 6 – Numărul de cuvinte aferent fiecărui nivel, pentru limbi diferite

Autori	A1
Van Ek & Alexander, 1980	
Van Ek, 1976	
Meara & Milton, 2003	<1,500
Beacco a.o. 2004	1,000
Rolland & Picoche, 2008	3,357

<sup>&</sup>lt;sup>541</sup> I.S.P. Nation, op. cit., 2016, p. 122.

<sup>&</sup>lt;sup>542</sup> Desigur că un individ care dorește să învețe româna pe teritoriul României va avea nevoie de unele elemente lexicale necesare pentru a putea "supraviețui" aici, în vreme ce o altă persoană care învață limba într-o altă țară va putea urma întocmai parcursul descris în programe.

<sup>&</sup>lt;sup>543</sup> Françoise Kusseling, Wilfried Decoo, Europe and language learning: The challenges of comparable assessment în 34th European Studies Conference – University of Nebraska – Omaha, Omaha, University of Nebraska, 2009, p. 7.

Milton, 2006	(400)
Institutul Cervantes, 2006	1,300
Bergan, 2001	
Franceză <sup>544</sup>	1160
Greacă <sup>545</sup>	1486

Coroborând informațiile, obținem următoarele valori (mult prea diferite) pentru nivelul A1: 400-1500 (3357) de elemente.

Întrucât rezultatele sunt atât de diferite, se poate folosi formula PERT<sup>546</sup> pentru a estima valoarea medie a acestor intervale: min + max + 4\*media / 6.

Așadar, utilizând această formula, valorile devin: A1: 1485.

**b.** De asemenea, vom încerca să aflăm numărul de ore folosind cealaltă posibilitate, a relației dintre numărul de cuvinte predate pe oră și numărul de ore pe nivel.

Platon<sup>547</sup> propune un număr aproximativ de ore necesare pentru însușirea structurilor aferente fiecărui nivel. Astfel, pentru A1 sunt necesare 80-100 de ore. Pe de altă parte, Gairns și Redman sugerează introducerea a 8-12 cuvinte pe oră (8 cuvinte, pentru începători, și 12, pentru avansați<sup>548</sup>). Și Nation și Waring susțin că este nevoie de aproximativ șapte expuneri la un cuvânt pentru a-1 putea considera ca însușit<sup>549</sup>. Având aceste cifre, putem încerca să stabilim un număr de cuvinte pentru fiecare nivel și din această perspectivă:

A1: 640 – 1200 (media: 920).

În urma rezultatelor obținute, vom încerca să ne situăm între cele două valori (a. și b.), având în vedere că diferențele dintre rezultatele obținute nu sunt foarte mari. Desigur că numărul exact de cuvinte pentru fiecare nivel nu poate fi estimat înainte de a avea inventarul, însă proiectarea unei limite ideale între niveluri este utilă în acest moment.

#### 5.6. Concluzii

În acest capitol am arătat importanța pe care o are vocabularul în procesul comunicării, în general, și am evidențiat principalele accepțiuni care au existat de-a

.

<sup>&</sup>lt;sup>544</sup> James Milton, Tomai Alexiou, *Developing a vocabulary size test in Greek as a foreign language* (prezentare ppt).

<sup>&</sup>lt;sup>545</sup> Ibidem.

<sup>&</sup>lt;sup>546</sup> Această formulă este folosită, în general, la estimarea orelor alocate unor activități în management. http://www.brighthubpm.com/methods-strategies/15682-introduction-to-the-pert-formula-series/

<sup>&</sup>lt;sup>547</sup> Elena Platon, *op. cit.*, 2009, pp. 499-505.

<sup>&</sup>lt;sup>548</sup> Ruth Gairns și Stuart Redman, *op.cit.*, 1986, p. 66.

<sup>&</sup>lt;sup>549</sup> Rob Waring, Paul Nation, op. cit., 2004, pp. 11-23.

lungul timpului cu privire la învățarea lexicală în cazul limbilor străine. Am continuat prin a prezenta care sunt principiile de urmat în elaborarea inventarelor lexicale pe baza LC, precum și criteriile folosite de specialiști în conceperea listelor de cuvinte. S-au evidențiat două direcții de urmat în elaborarea acestor inventare. Pe de-o parte, folosirea celor trei criterii arhicunoscute, frecvența, răspândirea și dispersia, iar, pe de altă parte, completarea acestora cu alte criterii, cu caracter subiectiv, în scopul obținerii unei liste cât mai relevante. Printre aceste criterii, se numără criteriul necesității și cel care vizează întregirea sferelor lexicale, pe care ne propunem să le folosim și noi în analiza care urmează. Întrucât considerăm foarte important să stabilim niște așteptări privind dimensiunea listelor de cuvinte rezultate în urma cercetării noastre, am încercat să stabilim în acest capitol limitele între care ne propunem să se regăsească numărul de cuvinte din inventarul lexical, aferent nivelului A1. Continuăm demersul nostru, în capitolul următor, cu prezentarea principalelor metode de analiză statistică a LC, pe de-o parte, iar pe de altă parte, prezentăm corpusul de producții scrise ale vorbitorilor nonnativi, Nivelul A1.

# CAPITOLUL VI – ROLUL STATISTICII ÎN LINGVISTICA BAZATĂ PE CORPUS (LC)

## 6.1. Statistica și LC

După cum am arătat în capitolul al patrulea, LC reprezintă o metodă cantitativă, ceea ce înseamnă că "lucrează cu numere, care reflectă frecvențele cuvintelor și ale expresiilor în corpusuri" [tr. n.]<sup>550</sup>. Datorită acestei trăsături, LC are nevoie de statistică pentru a putea raporta rezultatele într-o formă corectă și profesionistă. Din multitudinea definițiilor date *statisticii*, în cazul nostru se potrivește definirea statisticii drept "disciplina care ne ajută să înțelegem datele cantitative"<sup>551</sup>, adică, "știința colectării și a interpretării datelor care pot fi măsurate, cuantificate"<sup>552</sup>. Așa cum arată Brezina, ecuațiile și expresiile matematice reprezintă instrumente frecvent utilizate în statistică pentru a manipula datele pe parcursul analizei și pentru a le prezenta într-o formă mai accesibilă cititorului. Și în cazul LC sunt utilizate aceste ecuații sau expresii matematice<sup>553</sup>. Un prim exemplu, în acest sens, este reprezentat de **calculul mediei aritmetice** în scopul descrierii mostrei analizate, de pildă. O altă situație unde se utilizează expresii matematice este cea legată de exprimarea relației dintre două variabile care descriu mostra, obținând așanumita **regresie liniară** (*regression line* sau *line of the best fit*)<sup>554</sup>.

Şi, întrucât nu putem analiza şi prezenta rezultatele studiului nostru fără a face uz de statistică, definim în continuare termenii fundamentali din LC cu care vom opera, așa cum sunt descriși de Brezina și de către specialiștii în LC de la Universitatea Lancaster din Marea Britanie<sup>555</sup>. Am ales să traducem conceptele în limba română și să păstrăm în paranteză denumirile originale în limba engleză pentru a ne asigura că nu se pierde nimic din acuratețea termenilor. După conceptul principal vehiculat în lucrarea noastră, acela de **corpus**, putem vorbi despre **baza de date** (*dataset*). Aceasta reprezintă "o serie de constatări/descoperiri obținute din

<sup>&</sup>lt;sup>550</sup> Tony McEnery, Andrew Hardie, op. cit., 2011.

<sup>&</sup>lt;sup>551</sup> Vaclav Brezina, op. cit., 2018, p. 3.

<sup>&</sup>lt;sup>552</sup> P. J. Diggle, A. G. Chetwynd, *Statistics and scientific method: an introduction for students and researchers*, Oxford University Press, 2011, p. VII *apud* Vaclav Brezina, *op. cit.*, 2018, p. 3.

<sup>&</sup>lt;sup>553</sup> Vaclav Brezina, *op. cit.*, 2018, p. 3.

<sup>&</sup>lt;sup>554</sup> Pentru mai multe detalii legate de aceste două concepte, vezi, Vaclav Brezina, op. cit., 2018, pp. 3-4.

<sup>&</sup>lt;sup>555</sup> Vaclay Brezina, op. cit., 2018, p. 5-14.

corpus, care pot fi analizate" [tr. n.]<sup>556</sup> cu instrumente lingvistice în scopul obținerii unor concluzii pentru cercetarea întreprinsă. Baza de date constituie, de fapt, "o serie de rezultate individuale colectate din corpus, care sunt păstrate sub forma unui document *Excel*, de pildă, unde, fiecare linie reprezintă **un caz individual** (*data point/case*)" [tr. n.]<sup>557</sup>, adică un element lexical, în cazul nostru, iar "fiecare coloană reprezintă **o variabilă** (*variable*) distinctă" [tr. n.]<sup>558</sup>. După cum se înțelege din chiar denumirea sa, "prin **variabilă** înțelegem ceva ce poate varia sau poate avea valori diferite. De exemplu, vârsta subiecților poate constitui o variabilă, putând fi cuprinsă, în general, între 1 și 100 de ani sau vizând intervale mai mici, în funcție de tema de cercetare" [tr. n.]<sup>559</sup>.

Potrivit lui Brezina, de cele mai multe ori, cercetarea corpusului presupune căutarea unor variabile în interiorul corpusurilor, urmată de analiza relației dintre acestea. În cazul acestui concept, se distinge dihotomia variabile lingvistice (linguistic variables) vs variabile explicative (explanatory variables)<sup>560</sup>. Variabilele lingvistice "surprind frecventele anumitor trăsături lingvistice căutate în corpus" [tr. n.]<sup>561</sup>, în vreme ce variabilele explicative (numite si individuale – individual variables) "surprind contextul în care apar trăsăturile lingvistice" [tr. n.]<sup>562</sup>. Conform lui Brezina, o variabilă explicativă poate fi, spre exemplu, genul, vârsta, nivelul de competență lingvistică a subiectului sau anul de publicare și apartenența la un anumit gen literar, în cazul unui corpus care contine texte literare. Ambele tipuri de variabile (i.e. lingvistice și explicative) pot fi, la rândul lor, nominale, ordinale sau scalare/numerice<sup>563</sup> (nominal, ordinal, scale variables<sup>564</sup>). După cum arată Brezina, variabilele nominale reprezintă diferite categorii în care pot fi grupate cazurile din baza de date. Genul subiectilor, de exemplu, constituie o variabilă nominală, întrucât va fi necesară o selecție între a. genul masculin (care va primi un cod: de obicei, 0/M) și **b.** genul feminin (care va primi un cod: de obicei, 1/F)<sup>565</sup>. Variabilele ordinale presupun și o ierarhizare. "De exemplu, în cazul competenței lingvistice a subiecților

<sup>&</sup>lt;sup>556</sup> Vaclav Brezina, op. cit., 2018, p. 6.

 $<sup>^{557}</sup>$  Ibidem.

<sup>558</sup> Ibidem.

<sup>559</sup> Ibidem.

 $<sup>^{560}</sup>$  Ibidem.

<sup>&</sup>lt;sup>561</sup> Ibidem.

<sup>&</sup>lt;sup>562</sup> Ibidem.

<sup>&</sup>lt;sup>563</sup> Nicolae Valentin, Variabile statistice (clasificare și indicatori), Universitatea București, prezentare ppt, disponibilă pe: https://www.academia.edu/7577374/Variabile statistice

<sup>&</sup>lt;sup>564</sup> În LC, variabila scalară include atât variabila de interval, cât și variabila de raport.

<sup>&</sup>lt;sup>565</sup> După Vaclav Brezina, op. cit., 2018, p. 7.

într-o limbă, se poate face o ierarhie de la vorbitori începători la avansați" [tr. n.]<sup>566</sup>, respectiv, de la A1 la C2. Și în situația aceasta este indicat să folosim un sistem de etichetare a fiecărui nivel din ierarhia noastră (1 =începător, 2 =intermediar, 3 =avansat sau 1 = A1, 2 = A2, 3 = B1, 4 = B2, 5 = C1, 6 = C2). Variabilele scalare se diferențiază de celelalte două prin faptul că "pot avea orice valoare dintr-o scală ce arată "cantitatea" unei trăsături" [tr. n.]<sup>567</sup>.

Distributia frecventei (the frequency distribution) "ne oferă informatii privind valorile pe care le are o variabilă, precum și informații despre frecvența sa. O reprezentare grafică des utilizată pentru a arăta distribuția variabilelor scalare o contituie histograma" [tr. n.]<sup>568</sup>, iar o distribuiție uniformă<sup>569</sup> sau de referință (normal distribution) este reprezentată sub forma unui clopot simetric<sup>570</sup>. Specialistii arată că, în mod normal, "în natură se păstrează o distribuție uniformă, însă, atunci când analizăm date lingvistice, de cele mai multe ori, distribuția observată va fi descentrată/asimetrică" [tr. n.]<sup>571</sup>.Înainte de a accepta această asimetrie, singura conditie este să verificăm dacă, în datele noastre, nu se evidentiază o valoare extremă (outliner), "care se află foarte departe de celelalte valori obtinute" [tr. n.]<sup>572</sup> și care ne poate indica un rezultat corect, dar dificil de analizat. Sau, aceasta poate fi semnul apariției așa-numitei rogue value, adică a unei erori de măsurare, rezultată, spre exemplu, din introducerea gresită a datelor în baza de date, asa cum arată Brezina<sup>573</sup>. Dacă se dovedeste că avem doar o valoare extremă, este important să se ia o decizie care să nu îngreuneze analiza, existând chiar posibilitatea ignorării acestei valori<sup>574</sup>.

Măsurarea **mediei** (*average*) sau **a tendinței centrale** reprezintă, conform lui Brezina, un alt concept important în statistică, care apare, în general, împreună cu **dispersia** (*dispersion*) pentru a face o prezentare succintă a datelor. Există diferite tipuri ale mediei, iar, în cazul nostru, în LC, se folosesc **media aritmetică** (*mean*) și **mediana** (*median*). Media aritmetică este reprezentată de "suma tuturor valorilor, împărțită la numărul cazurilor" [tr. n.]<sup>575</sup>. Brezina recomandă utilizarea acesteia doar în situațiile în care nu avem valori extreme. Dacă în analiza noastră obținem astfel

<sup>&</sup>lt;sup>566</sup> Vaclav Brezina, op. cit., 2018, p. 7.

 $<sup>^{567}</sup>$  Ibidem.

<sup>&</sup>lt;sup>568</sup> *Ibidem*, p. 8.

<sup>&</sup>lt;sup>569</sup> Nicolae Valentin, op. cit.

<sup>&</sup>lt;sup>570</sup> Vaclav Brezina, op. cit., 2018, p. 8.

<sup>&</sup>lt;sup>571</sup> Ibidem.

<sup>&</sup>lt;sup>572</sup> Ibidem.

<sup>&</sup>lt;sup>573</sup> Ibidem.

<sup>&</sup>lt;sup>574</sup> *Ibidem*, pp. 9-10.

<sup>&</sup>lt;sup>575</sup> *Ibidem*, p. 10.

de rezultate, este indicat să folosim mediana, care e definită, "în cazul unui șir impar, drept valoarea centrală, iar, în cazul unui șir par, este reprezentată de media aritmetică a celor două valori din centrul șirului" [tr. n.]<sup>576</sup>. E bine de respectat această recomandare pentru a evita niște rezultate prea tare influențate de valorile extreme.

Un alt concept important în statistica utilizată în LC este acela de **dispersie** (*dispersion*), calculată drept "răspândirea valorilor unei variabile într-o bază de date" [tr. n.]<sup>577</sup>, mai exact, aceasta ne arată cât de departe se găsesc valorile una de cealaltă. În statistică, distanța de la o valoare la cealaltă se numește **împrăștiere**<sup>578</sup> (range<sub>1</sub>), diferită de **răspândire** (range<sub>2</sub>), conceptul din LC. Întrucât împrăștierea poate fi influențată de valoarea extremă, Brezina recomandă să calculăm **amplitudinea** (*interquartile range*) sau să folosim ecuația de calcul a **abaterii standard** (*standard deviation*). Toate aceste modalități de calcul ale dispersiei pot fi făcute cu ajutorul unei foi de calcul, de tipul *Excel*<sup>579</sup>.

La fel ca Nation, și Brezina susține că nu există un răspuns universal referitor la cât de mare ar trebui să fie un corpus pentru a obține rezultate relevante, deoarece dimensiunea corpusului depinde și de tema cercetării. De pildă, dacă se doresc informații despre subiecte comune, atunci corpusul poate avea o dimensiune relativ restrânsă, de aproximativ un milion de cuvinte sau chiar mai puține. Dacă, în schimb, dorim să identificăm răspândirea anumitor elemente lexicale, dar și a expresiilor din care fac parte acestea, va fi nevoie de un corpus mult mai extins, deoarece unele cuvinte sunt destul de rare în producțiile indivizilor<sup>580</sup>.

Acesta este și unul dintre motivele pentru care am hotărât să ne restrângem cercetarea la identificarea elementelor lexicale de la nivelul A1 (care, pe baza experienței didactice, presupunem că nu sunt extrem de numeroase), cu toate că, în propunerea inițială, am intenționat să elaborăm inventarele lexicale pentru nivelurile A1-B2. Desigur că această limitare a cercetării nu înseamnă că nu este necesar un demers similar și în cazul nivelurilor superioare, ci presupune doar o amânare a acestei analize, pentru momentul în care vom dispune de un corpus mai extins și, ideal, compus și din producții ale studenților din alte centre universitare pentru a ne asigura de obținerea unor rezultate cât mai relevante.

<sup>&</sup>lt;sup>576</sup> Vaclav Brezina, op. cit., 2018, pp. 10-11.

<sup>&</sup>lt;sup>577</sup> *Ibidem*, p. 11.

<sup>&</sup>lt;sup>578</sup> Nicolae Valentin, *op. cit.* 

<sup>&</sup>lt;sup>579</sup> Vaclav Brezina, op. cit., 2018, p. 11.

<sup>&</sup>lt;sup>580</sup> *Ibidem*, pp. 18-19.

Prin urmare, obiectul de studiu al LC este reprezentat de corpus, format din cuvinte și expresii, adică din variabile lingvistice, așa cum am arătat mai sus. "De obicei, în LC numărăm cuvintele în contexte diferite de utilizare și apoi comparăm aceste analize pentru a identifica anumite **tipare de utilizare a limbii** (patterns of language use)" [tr. n.]<sup>581</sup>. După procesul de numărare a cuvintelor, vom obține un număr de ocurențe în corpus, o valoare ce ne arată frecvența cuvântului, cunoscută drept **frecvența absolută/brută** – **FA** – (absolute/raw frequency) a unui element lexical. Conform lui Brezina, FA este cel mai des utilizată în cazul analizei unui singur corpus și în scopul obținerii unei liste ierarhizate a elementelor lexicale din acel corpus. Pentru analize mai complexe sau pentru compararea a două sau a mai multor corpusuri, se folosește **frecvența relativă** – **FR** – (relative/normalized frequency), care se calculează ținând cont de FA și de numărul de elemente din corpus, folosind formula<sup>582</sup>:

$$FR = \frac{FA}{num "ar" elemente corpus} \times baza de normalizare^{583}$$

După cum arată Brezina, baza de normalizare se stabilește în funcție de dimensiunile corpusului. Dacă avem de-a face cu un corpus extins, baza de normalizare folosită este de 1.000.000. Iar, cu cât e mai redus corpusul, cu atât baza de normalizare va fi mai restrânsă, 10.000, 1.000 sau chiar și 100, în cazul unui corpus de mici dimensiuni. Baza de normalizare este necesară pentru a putea raporta FR, adică pentru a găsi o variantă "de prezentare a datelor mai simplu de gestionat de către cititorul rezultatelor noastre" [tr. n.]<sup>584</sup>. În mod indiscutabil, este mai simplu de analizat, de reținut sau de prelucrat baza de date atunci când se manevrează valori de ordinul miilor decât de ordinul milioanelor, de pildă.

Principiul prin care se reduc cu rapiditate frecvențele cuvintelor într-un corpus este cunoscut drept legea lui Zipf. Potrivit lui Brezina, doar primul sau primele cuvinte dintr-o listă de frecvențe va/vor avea un număr de ocurențe foarte mare, după care procentul acestora va scădea, însă predictibil<sup>585</sup>.

137

<sup>&</sup>lt;sup>581</sup> Vaclav Brezina, *op. cit.*, 2018, p. 39.

<sup>&</sup>lt;sup>582</sup> *Ibidem*, pp. 42-43

<sup>&</sup>lt;sup>583</sup> Toate ecuațiile folosite în acest capitol sunt preluate și traduse din Vaclav Brezina, *op. cit.*, 2018, pp. 43-51.

<sup>&</sup>lt;sup>584</sup> Vaclav Brezina, *op. cit.*, 2018, p. 43.

<sup>&</sup>lt;sup>585</sup> *Ibidem*, p. 44.

Așadar, atunci când raportăm frecvența unui element din lista noastră, este important să includem atât frecvența absolută (FA), cât și frecvența relativă (FR), raportată la baza de normalizare, adică, de pildă: FA = 3.567.890, FR = 3.567/milion.

Legea lui Zipf stipulează că "există o relație inversă între frecvența cuvintelor și rangul lor din lista de frecvențe"<sup>586</sup>. Astfel se poate observa un model care arată gradul de regularitate cu care descrește frecvența cuvintelor în limba naturală: primele cele mai frecvente cuvinte se vor întâlni aproximativ de două ori mai des decât cuvintele aflate pe locul doi din punctul de vedere al frecvenței și de trei ori mai des decât cele aflate pe locul trei etc. <sup>587</sup>. În plus, într-o listă de frecvență a cuvintelor, dacă înmulțim rangul (numărul de ordine al cuvântului) cu frecvența sa, vom obține o valoare constantă (aceeași sau foarte apropiată) pe măsură ce înaintăm în listă <sup>588</sup>. Mai mult, bazându-se pe exemple din listele elaborate, precum și pe legea lui Zipf, Nation arată că jumătate dintre cuvintele dintr-o astfel de listă apar doar o dată în text, fiind identice cu cele mai reprezentative cuvinte din punctul de vedere al conținutului pentru textul sau colecția de texte la care se face referire <sup>589</sup>.

Tot pe baza analizelor empirice de vocabular, aplicate limbii engleze, se arată că, în general, cele mai frecvente zece cuvinte acoperă aproximativ 25% dintre elementele lexicale dintr-un text în limba engleză. Vorbim aici despre cuvinte precum: *the*, *and*, *a*, *of*, *to*, *s*, *as*. Această afirmație a fost verificată și în cazul limbii maori (populația indigenă din Noua Zeelandă), unde s-a arătat că primele 10 cuvinte ca frecvență acoperă 35% dintre elementele unui text. Alte cercetări, tot în cazul limbii engleze, ne arată că cele mai frecvente 100 de cuvinte din engleză acoperă aproximativ 50% dintre cuvintele unui text, iar, primele 1000 de cuvinte, între 70% și 90%, depinde de tipul textului (discurs oral sau scris), dar și de tema acestuia<sup>590</sup>. Va fi interesant de descoperit care este situația în cazul limbii române, cel mai probabil, principiul verificându-se și în cazul limbii noastre.

În scopul descrierii complete a unui cuvânt sau a unei expresii dintr-un corpus, pe lângă frecvență, se impune să analizăm și dispersia, pentru a evita așanumita *whelk problem* = problema melcului de mare<sup>591</sup>. Aceasta este atribuită lui Kilgariff și constituie o metaforă pentru a ilustra o situație posibilă și frecventă în LC, și anume distribuția inegală a cuvintelor într-un corpus. Pentru a ilustra această

<sup>&</sup>lt;sup>586</sup> Paweł Szudarski, op. cit., 2018, p. 55.

<sup>587</sup> Ibidem.

<sup>&</sup>lt;sup>588</sup> După I.S.P. Nation, op. cit., 2016, p. 3.

<sup>589</sup> Ibidem.

<sup>&</sup>lt;sup>590</sup> *Ibidem*, p. 4.

<sup>&</sup>lt;sup>591</sup> Vaclav Brezina, op. cit., 2018, p. 46.

teorie, Brezina oferă următorul exemplu: este posibil ca, în corpusul folosit, să existe un fragment substanțial dintr-o carte despre melcii de mare, de exemplu, însă, în restul materialului, să nu mai fie mostre în legătură cu această temă, situație care va avea consecințe asupra rezultatelor pe care le vom obține. Urmarea imediată va fi o frecvență ridicată (mult mai mare decât realitatea) a acestui cuvânt (*i.e.* melc de mare), ceea ce va conduce spre rezultate eronate ale analizei noastre<sup>592</sup>. De aceea e important ca distribuția să evidențieze prezența unui cuvânt/a unei structuri de-a lungul corpusului. Un alt aspect de reținut, conform lui Brezina, este acela că dispersia se poate măsura cu ajutorul mai multor metode, în funcție de corpus și de ceea ce dorim să analizăm. Cea mai cunoscută metodă este calculul răspândirii unui cuvânt în corpus (*range*<sub>2</sub>). Aceasta se calculează după formulele<sup>593</sup>:

răspândirea (R) = numărul părților din corpus unde apare cuvântul/expresia

sau  $r\"{a}sp\^{a}ndirea\% = \frac{num\~{a}rul\ p\~{a}r\~{t}ilor\ din\ corpus\ unde\ apare\ cuv\^{a}ntul/expresia}{num\~{a}rul\ total\ de\ p\~{a}r\~{t}i\ ale\ corpusului}} \times 100$ 

Totuși, specialiștii recomandă utilizarea acestei metode doar pentru o analiză primară a copusului, întrucât oferă un răspuns de tip *da/nu*, ceea ce poate constitui o problemă în situații în care avem cuvinte sau expresii cu aceleași frecvențe sau similare. În astfel de cazuri, ne poate spune doar că un cuvânt sau o expresie este prezent/ă într-un subcorpus sau într-un text, fără a raporta ocurențele în fiecare caz. Folosind această metodă, nu vom putea identifica și evita, de exemplu, *melcul de mare* <sup>594</sup>.

Calculul abaterii standard (AS)<sup>595</sup> reprezintă o altă metodă de calcul al dispersiei, mult mai cunoscută în statistică, în general. Aceasta arată cât de departe se află FR individuale ale cuvintelor față de media FR într-o bază de date. AS se calculează după formula<sup>596</sup>:

$$AS = \sqrt{\frac{suma\ pătratelor\ distanței\ față\ de\ media\ FR}{numărul\ total\ de\ părți\ ale\ corpusului}}$$

101uem.

<sup>&</sup>lt;sup>592</sup> Vaclav Brezina, op. cit. 2018, pp. 46-47.

<sup>&</sup>lt;sup>593</sup> Preluate din exemplele lui Vaclav Brezina, op. cit., 2018, p. 48.

<sup>&</sup>lt;sup>594</sup> Vaclav Brezina, *op. cit.*, 2018, p. 48.

<sup>&</sup>lt;sup>595</sup> Ibidem.

<sup>&</sup>lt;sup>596</sup> *Ibidem*, p. 49.

Cu toate că e cea mai cunoscută metodă de calcul al dispersiei, ea "nu este relevantă în cazul nostru (sau nu e relevantă de una singură), întrucât nu ne oferă posibilitatea comparării dispersiei cuvintelor care apar cu frecvente diferite" [tr. n.]<sup>597</sup>. În astfel de cazuri, Brezina recomandă să se folosească coeficientul de variatie - CV - (coefficient of variation) sau asa numitul D al lui Juilland (Juilland's D).

Coeficientul de variatie se calculează după formula:

$$CV = \frac{abaterea\ standard}{media}$$

Conform lui Brezina, D-ul lui Juilland constituie o metodă de măsurare a dispersiei, dezvoltată de Juilland pentru elaborarea dictionarelor de frecventă. Acesta este un număr între 0 și 1, unde 0 reprezintă o distribuție extrem de inegală, iar 1, o distribuție perfect egală. Se calculează astfel<sup>598</sup>:

$$D-ul\ lui\ Juilland = 1 - \frac{coeficientul\ de\ variație}{\sqrt{numărul\ părților\ corpusului-1}}$$

# 6.2. Corpusul de productii scrise ale vorbitorilor nonnativi de limba română ca limbă străină. Nivelul A1

Înainte de a elabora un nou corpus, se impune verificarea, printr-un set de criterii ferme, a necesității și a relevanței acestuia. Nelson propune un set de întrebări<sup>599</sup> în vederea dovedirii necesității noului corpus, pe care le prezentăm mai jos:

1. Acest corpus e necesar? Înainte de a concepe corpusul, se recomandă să verificăm dacă vreunul dintre corpusurile deja existente nu ar putea răspunde nevoilor noastre. Răspunsul la această întrebare l-am formulat deja în câteva rânduri în lucrarea noastră. La momentul redactării lucrării, nu aveam cunoștință să existe un corpus publicat de producții scrise ale vorbitorilor nonnativi. Totuși, în cursul anului 2020, a apărut Româna ca limbă străină. Corpus<sup>600</sup>, care include

<sup>&</sup>lt;sup>597</sup> Vaclav Brezina, op. cit. 2018, p. 50.

<sup>&</sup>lt;sup>598</sup> *Ibidem*, p. 51.

<sup>&</sup>lt;sup>599</sup> Mike Nelson, "Building a written corpus. What are the basics?", în Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012, p. 55.

<sup>&</sup>lt;sup>600</sup> Este vorba despre: Mihaela-Viorica Constantinescu, Gabriela Stoica, op. cit., 2020.

si productii scrise, si productii orale ale vorbitorilor nonnativi, dar care contine productii de pe întreg parcursul Anului pregătitor. Pe de altă parte, corpusurile electronice, disponibile pentru RLM, nu considerăm că pot fi reprezentative pentru cercetarea noastră sau, cel puțin, nu pot fi folosite singure, ci împreună cu un corpus specializat, al vorbitorilor nonnativi. Amintim că, în cercetarea de fată, vom folosi si CORLS, care, în acest moment, constituie singurul corpus disponibil de productii orale ale vorbitorilor nonnativi.

- 2. Cât de extins ar trebui să fie? Uneori, se impune un corpus extins, alteori, unul redus se dovedeste a fi suficient. Nelson sustine că scopul pentru care e folosit corpusul dictează dimensiunea acestuia. Sinclair și-a bazat teoriile privind necesitatea unui corpus cât mai mare pe ideea că în text cuvintele nu sunt distribuite în mod egal, putându-se întâmpla ca unele cuvinte să apară numai o dată. "Astfel, pentru a studia comportamentul cuvintelor în text, noi trebuie să avem suficient de multe ocurente" [tr.n.]601.
- 3. Corpusul reprezintă un esantion echilibrat, dar de referintă pentru limbă? În ceea ce priveste productiile incluse în corpus, în general, specialistii au păreri împărțite – unii sugerează fragmente de 2.000 de cuvinte, alții, de 20.000, alții, materialul integral<sup>602</sup>. În cazul nostru, din rațiuni de disponibilitate a materialelor, precum și din cauza numărului redus de cuvinte care compun o producție la nivelurile A1 și A2, am optat pentru întregul material, de fiecare dată.
- 4. Cum întocmim corpusul? Poate cea mai complicată etapă este aceea de adunare a materialului și de introducere în format electronic. Dacă textele nu sunt în format electronic, acestea vor trebui introduse în acest format, iar dacă acestea reprezintă productii ale elevilor sau ale copiilor, "este important să se păstreze redactarea nonstandard și structurile gramaticale folosite inițial, chiar dacă sunt incorecte, deoarece vor fi extrem de relevante. Decizii vor trebui luate și în ceea ce priveste diferite semne grafice folosite" [tr. n.]<sup>603</sup>.

Un alt aspect extrem de important în întocmirea corpusului este acela legat de drepturile de autor, precum și aspectele legate de confidențialitate. Întrucât folosim doar producțiile studenților din cadrul programelor de studiu de An pregătitor și Erasmus de la Facultatea de Litere, Universitatea Babeș-Bolyai, am transcris în format electronic doar textele studenților care și-au exprimat acordul ca

<sup>601</sup> Sinclair apud Mike Nelson, op. cit., 2012, p. 55.

<sup>&</sup>lt;sup>602</sup> *Ibidem*, p. 58.

<sup>603</sup> Randi Reppen, "Building a corpus. What are the key considerations?", în Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012, p. 35.

producțiile lor să fie folosite. În plus, orice referire la numele autorului unei producții a fost înlăturată<sup>604</sup>.

În ceea ce priveste adnotarea, Sinclair recomandă să avem o variantă a corpusului "curată", fără adnotări, și una cu adnotări, deoarece nu toți cercetătorii folosesc în același fel și în același scop corpusul, iar noi nu îl realizăm numai pentru uzul nostru. Deocamdată, am decis să nu adnotăm corpusul, întrucât această procedură se face în mare parte automat, considerând că ar fi mai bine să se extindă pe viitor materialul și abia după aceea să fie adnotat. Mai mult, pentru cercetarea noastră nici nu avem nevoie de un corpus adnotat, întrucât nu ne interesează ordinea structurilor gramaticale, ci a elementelor lexicale.

Pentru cazurile în care avem în corpus elemente nedorite sau cel putin nedorite pentru o anumită analiză, recomandarea este ca acestea să fie izolate între paranteze unghiulare (<x>)"605.

O altă decizie care trebuie luată este aceea cu privire la condițiile în care a fost produsă mostra de limbă. Aici vorbim despre două conditii. Una se referă la timpul de redactare. A fost contorizat sau nu? Iar cea de-a doua se raportează la materialele de referință, precum dictionare sau cărți de gramatică, de care se pot folosi studenții, în timp ce scriu. În funcție de ceea ce urmăresc autorii corpusului, producțiile studenților pot fi spontane sau atent lucrate, caz în care vor avea acces la materiale de referintă. "ICLE demonstrează că o sursă bună de productii scrise o reprezintă eseurile redactate în cadrul evaluărilor"606. Unii ar putea sustine că aceste mostre de limbă nu reprezintă o limbă naturală, însă Ellis și Barkhuizen contrazic această afirmație, întrucât "examinarea reprezintă pentru studenți un context natural în care îsi pot exersa L2"607. Mai mult, Constantinescu si Stoica consideră că:

"Temele redactate în cadrul testelor și examenelor scrise sunt, poate, cele mai relevante pentru nivelul de achizitie a limbii (studentul nu are posibilitatea corectării prin utilizarea unor materiale suport, sarcina trebuie rezolvată într-un anume interval de timp, studentul este conștient de evaluarea riguroasă a textului produs și deci este, în principiu, mai atent în redactarea lui). 608

<sup>604</sup> La transcrierea variantei brute a producțiilor au participat și studenți masteranzi de la modulul de RLS, dar si studenti, de la nivel licentă, coordonati de către lect, univ, dr. Dina Vîlcu.

<sup>605</sup> I.S.P. Nation, op. cit., 2016, p. 108.

<sup>606</sup> Rod Ellis, Gary Barkhuizen, op. cit., 2005, p. 30.

<sup>&</sup>lt;sup>607</sup> *Ibidem*, p. 50.

<sup>608</sup> Mihaela-Viorica Constantinescu, Gabriela Stoica, Româna ca limbă străină. Corpus, Editura Universității București, București, 2020, p. 25.

În elaborarea corpusului nostru, am ales să folosim producțiile scrise ale studenților în cadru organizat, în timpul examenelor de final de nivel, întrucât, din punctul nostru de vedere, reprezintă cel mai bine limba însusită de student la sfârsitul unei etape de învățare și ne pot releva mult mai multe despre bagajul lexical al unui vorbitor nonnativ la nivelul A1. Am luat în considerare si utilizarea unor texte elaborate pe parcursul învățării, însă credem că ar fi existat pericolul reprezentării în productiile studentilor a unui lexic neînsusit sau însusit insuficient<sup>609</sup>, ceea ce ar fi condus la rezultate irelevante. Cu toate că, în timpul examenului, nu se contorizează timpul de redactare a productiei scrise, de-a lungul anilor 2013-2017, la Departamentul de limbă, cultură si civilizatie românească al UBB s-au efectuat mai multe analize pentru a stabili durata necesară rezolvării în bune conditii a întregului examen<sup>610</sup>. Astfel, putem spune că, deși timpul de redactare nu este contorizat pentru producțiile scrise, poate fi evidențiat un interval aproximativ necesar pentru redactarea subiectului de producere scrisă. Asadar, s-a arătat că, pentru A1, studentii au nevoie de circa 30 de minute (27',93" - 33', 64") pentru a redacta cele două texte<sup>611</sup>.

# 6.3. Descrierea corpusului de producții scrise ale vorbitorilor nonnativi, nivelul A1

Corpusul elaborat în prezenta lucrare constituie un corpus de limbă scrisă, diacronic, specializat, pentru limba română ca L2, nivelul A1. Așadar, un material destinat unei limbi de circulație restrânsă, pentru un nivel elementar.

Corpusul conține un număr de 799 de producții scrise (texte) de română ca L2, aparținând unui număr de 451 de studenți din programele universitare de An pregătitor (AP) și Erasmus (E), de la Facultatea de Litere a Universității Babeș-Bolyai. Acestea au fost colectate în cadrul sesiunilor de examinare de nivel A1, în perioada: 2013-2020 (AP), respectiv, 2014-2017 (E).

#### 6.3.1 Subiecții

Pentru a avea la dispoziție toate elementele necesare unei analize cât mai complexe și relevante, înainte de fiecare sesiune de examinare, s-au colectat, de la

143

<sup>609</sup> Pe de-o parte, din cauza faptului că studenții nu au parcurs un nivel complet, iar, pe de altă parte, din cauza accesului la alte resurse, de tipul dictionarelor sau al altor materiale ajutătoare.

<sup>&</sup>lt;sup>610</sup> Despre procesul de stabilire a timpilor pentru toate examenele de competență în limba română ca limbă străină, vezi Antonela Ariesan, *op. cit.*, 2017, p. 135-152.

<sup>611</sup> Antonela Ariesan, op. cit., 2017, p. 143, 145.

fiecare student, date cu privire la: L1, alte limbi străine cunoscute, sex, vârstă, țara de proveniență, perioada de instruire în L2, studii și motivele pentru care au ales să învete în România.

În ceea ce privește L1<sup>612</sup>, putem observa că avem de-a face cu un grup de subiecți ce provin din medii culturale diferite, așadar, și cu L1 diverse. În analiza L1, am decis să grupăm sub *L1 asiatice* limbile: chineza, japoneză, coreeană, filipineză, tailandeză și vietnameză, iar, sub denumirea de *L1 africane*, limbile studenților din Nigeria, Camerun și Kenya, exceptând țările cu limba oficială arabă: Maroc, Egipt, Algeria, Tunisia, Libia, Sudan.

<sup>&</sup>lt;sup>612</sup> O situație aparte o pot reprezenta studenții care provin din țări în care limba maternă nu coincide întotdeauna cu limba de instrucție. De exemplu, studenții din Israel și cei din Maroc, care au limbile de studiu ebraica, respectiv, franceza. În acest caz, am ales să nu îi considerăm bilingvi, întrucât atât franceza, cât și ebraica reprezintă limbi de studiu, învățate mai târziu, în momentul începerii instruirii scolare și deci nu reprezintă limba dominantă, cea care reflectă identitatea culturală.

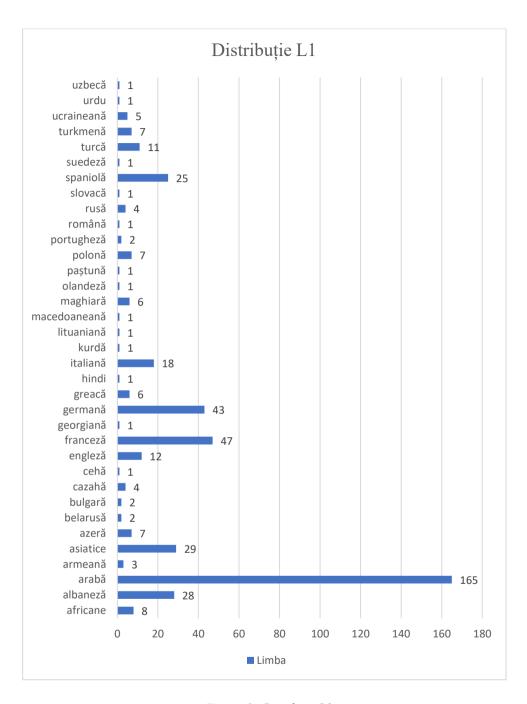


Figura 6 – Distribuția L1

După cum se observă din graficul de mai sus, distribuția L1 nu este una echilibrată, remarcându-se preponderența vorbitorilor de arabă ca L1, ceea ce, din

punctul nostru de vedere, caracterizează grupul-țintă, iar corpusul este, așadar, reprezentativ pentru această categorie de subiecți. Este de reținut că, în general, în AP se remarcă preponderența vorbitorilor de arabă.

Observăm că o reprezentare bună între subiecții noștri sunt cei care au L1 franceză și germană, acest lucru datorându-se, în opinia noastră, în primul rând, includerii studenților Erasmus în analiză. Având în vedere ponderea mare a acestor studenți în grupul nostru, alături de cei având spaniola ca L1, putem afirma că rezultatele noastre nu sunt relevante doar pentru studenții din AP, majoritar arabi.

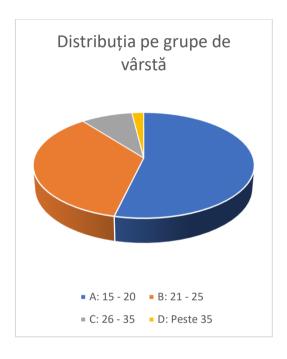


Figura 7 - Distribuția pe grupe de vârstă

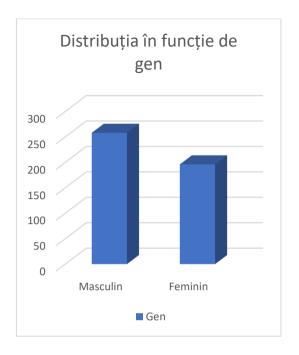


Figura 8 - Distribuția în funcție de genuri

În analiza noastră privind subiecții, am avut în vedere următoarele grupe de vârstă: A: 15-20 de ani, B: 21-25 de ani, C: 26-35 de ani, D: peste 35 de ani. După cum se observă din graficele de mai sus, există o preponderență a genului masculin și a grupei de vârstă A (15-20 de ani). Pe de altă parte, se remarcă o reprezentare mai slabă a categoriei D, de peste 35 de ani.

În ceea ce privește alte limbi străine cunoscute, 73% vorbesc și engleza la un nivel avansat sau intermediar, iar aproximativ 36% sunt vorbitori de limbi romanice, la nivel avansat sau intermediar<sup>613</sup>.

Dintre cei 452 de studenți, numai doi s-au înscris din motive personale la cursurile de RLS. Ceilalți 450 au drept motivație continuarea studiilor. Toți subiecții din AP au parcurs șase săptămâni de cursuri intensive de limba română (a câte 5 ore pe zi de RLS), în vreme ce studenții Erasmus au urmat un curs intensiv de 4 ore pe săptămână, de-a lungul unui semestru<sup>614</sup>. Deși, la prima vedere, se pare că avem de-

614 Numărul mai mic de ore se va observa, de cele mai multe ori, în lungimea și complexitatea textelor.

<sup>613</sup> Întrucât nu avem cum să verificăm cele declarate cu privire la nivelul de cunoaștere al altor limbi străine, am decis să nu luăm în calcul și nivelul începător, care, din experiența noastră cu studenții, poate fi, mai degrabă un preA1.

a face cu două grupuri-țintă extrem de diferite (AP și E)<sup>615</sup>, considerăm că, prin tratarea producțiilor lor împreună, ne asigurăm de o acoperire mai bună a grupului-țintă. Din punctul nostru de vedere, dacă ne-am fi limitat colectarea datelor doar de la studenții din AP, am fi riscat obținerea unui corpus și, implicit, a unui inventar lexical mult prea restrâns, deloc reprezentativ pentru româna predată altor categorii de vorbitori de L1. Desigur că, în continuare, pentru un grad și mai ridicat de reprezentativitate a celor două instrumente propuse prin lucrarea noastră, ar fi nevoie de o extindere a numărului producțiilor și a centrelor universitare de la care se face colectarea.

#### 6.3.2. Producțiile

Textele colectate pentru corpusul de producții scrise ale vorbitorilor nonnativi, de nivel A1, au fost transcrise în format electronic respectând, în cazul fiecărei producții, întocmai așezarea în pagină și modalitatea de redactare pentru care au optat studenții. Astfel, se vor putea remarca, în cadrul corpusului, texte scrise doar cu majuscule sau texte cu așezare deficitară în pagină. Am ales să păstrăm stilul și să nu intervenim în niciun fel, deoarece dorim ca acest corpus să fie cât mai ilustrativ pentru nivelul A1 și să constituie baza pentru alte studii legate de achiziția limbii române ca limbă străină, și nu numai.

Considerăm că propunem un corpus *reprezentativ pentru nivelul A1*, întrucât toate producțiile au fost colectate pe parcursul examenului de RLS, nivelul A1, examen ce face parte din bateria de teste acreditate ALTE. Apartenența la acest set de teste garantează că sarcinile propuse la scriere aparțin nivelului declarat și presupun același nivel de dificultate. Tot datorită acreditării ALTE și atelierelor de evaluare a producțiilor orale și scrise ce se desfășoară periodic la Departamentul de limbă, cultură și civilizație românească, putem afirma că evaluarea este cât se poate de obiectivă, așteptările fiind corelate cu cunoștințele predate studenților, iar sarcinile pe care le-au primit studenții, de-a lungul anilor, presupun aceeași structură, dar și același nivel de dificultate.

În procesul de formare a corpusului, am întâmpinat unele probleme cu privire la variabilele explicative în cazul subiecților, ceea ce a făcut ca în această situație să fim nevoiți să renunțăm la unele producții. De asemenea, textele extrem

148

\_

<sup>615</sup> Se remarcă ritmul mai rapid în care se parcurge materia la Erasmus, spre deosebire de AP, întrucât, de multe ori, studenții arabi din AP (o parte dominantă a grupului-țintă) nu sunt suficient de familiarizați cu alfabetul latin.

de scurte, notate foarte slab la evaluare<sup>616</sup>, au fost eliminate, întrucât erau și neinteligibile, așadar, inutilizabile din orice perspectivă de analiză. De îndată ce o producție a conținut și elemente care pot fi ilustrative pentru achiziția L2, am decis să o păstrăm în corpus.

Secțiunea de scriere din cadrul Examenului de RLS, de nivel A1, este compusă din două sarcini (1, 2 – vezi ultima cifră din codul studenților din corpus), ambele cu caracter obligatoriu. Timpul aproximativ necesar elaborării întregii probe de scriere este de 30 de minute. Cele două texte reprezintă două tipuri diferite:

- 1. cu *tipar interacțional* și cu o dimensiune de circa 30-40 de cuvinte. Textele elaborate sunt carte poștală sau mesaj.
- 2. de tip narativ sau descriptiv (50-60 de cuvinte).

Conform documentului de specificații pentru creatorii de teste, de la Departamentul de limbă, cultură și civilizație românească, prin proba de scriere se vizează următoarele abilități: "capacitatea de a redacta și organiza adecvat texte scurte și simple, raportate la date și experiențe personale"<sup>617</sup>. De asemenea, în procesul de redactare a celor două texte, candidatul trebuie să țină cont de toate punctele de sprijin indicate în cerințe. Menționăm că, în cazul studenților Erasmus, nu există de fiecare dată colectate ambele texte, întrucât aceștia, la unele examene, au avut doar un text de elaborat.

În tabelul de mai jos prezentăm tipurile de texte și temele conținute în corpusul nostru:

Tabelul 7 – Tipurile de texte și de teme din corpusul de producții scrise ale vorbitorilor nonnativi de RLS, Nivelul A1

Sarcina	Abilitatea verificată	Tipul de text	Numărul de	Teme reprezentate în
			cuvinte	corpus
1	capacitatea de a redacta	cu tipar	30-40 de	Școala (271)
	și de a organiza adecvat	interacțional	cuvinte	Vacanța (103)
	texte scurte și simple,	(mesaj sau carte		Timp liber (25)
	raportate la date și	poștală)		Activități zilnice (32)
2	experiențe personale	de tip descriptiv	50-60 de	Familia (271)
		sau narativ	cuvinte	Casa (148)
				Masa (29)

<sup>616</sup> Evaluarea producțiilor scrise se face de către doi evaluatori (Ev<sub>1</sub> și Ev<sub>2</sub>), pe baza unei grile cu patru criterii: complexitate, acuratețe, organizare și eficiența comunicării. În caz de diferențe mai mari de două puncte între punctajul acordat de Ev<sub>1</sub>și Ev<sub>2</sub>, se face supracorectură, acordându-se încă un punctaj. Punctajul maxim pe care un student îl poate obtine reprezintă 20 de puncte.

617 Specificații pentru creatorii de teste. Limba română ca limbă străină, nivelurile A1-B2, de la Departamentul de limbă, cultură și civilizație românească, Facultatea de Litere, Universitatea Babeș-Bolyai.

149

După cum se poate observa din *Tabelul 7*, temele de scriere pentru acest nivel nu sunt foarte diverse, întrucât la A1 ne raportăm doar la experiența personală a studenților. Pentru identificarea în corpus a temei, am încercat să găsim niște "etichete" care să ilustreze cât mai bine conținutul cuprins în tema respectivă. Prezentăm, pe larg, în *Tabelul 8*, temele, fără a detalia punctele de sprijin pe care le primesc studenții:

Tabelul 8 – Temele producțiilor scrise

Tema	Sarcina
Şcoala	mesaj, de tip e-mail, în care se prezintă o zi de școală
Vacanța	mesaj, de tip carte poștală, dintr-o destinație de vacanță
Timp liber	mesaj, de tip e-mail, în care se prezintă o zi liberă
Activități zilnice	mesaj, de tip e-mail/scrisoare, în care se prezintă activitățile zilnice
Familia	descriere de persoane
Casa	descrierea camerei
Masa	prezentarea meselor zilnice ale studentului

Am decis să considerăm teme diferite *Vacanța* și *Timpul liber*, deoarece și vocabularul utilizat la redactarea textelor pe cele două teme este diferit. De asemenea, în ceea ce privește tema *Familia*, cu sarcina de descriere a unei persoane, am decis să considerăm că este vorba despre familie și dacă studentul a descris un prieten. Dacă ne referim la ansamblul temelor, în general, am încercat să păstrăm un echilibru, pe cât a fost posibil, în ceea ce privește reprezentarea tuturor sferelor lexicale în corpus. Totuși, chiar dacă există unele teme cu producții mai multe, din conținutul acestora am putut observa că lexicul vehiculat se întinde și în afara temei textului, situație care, considerăm noi, este în favoarea noastră, arătându-ne cunoștințele lexicale deținute de studenți. Se vor observa situații în care, la prima vedere, unele texte nu par să aparțină de tema enunțată, deoarece studentul nu a respectat sarcina. Le-am păstrat așa, deoarece considerăm că, din perspectiva lexicului, sunt ilustrative și în acest fel, însă ele pot fi extrem de relevante și la o analiză a gradului de adecvare a textului la temă.

Așadar, corpusul de producții scrise ale vorbitorilor nonnativi de RLS, nivel A1, este organizat *tematic*, în subcorpusuri. În cadrul fiecărei teme/fiecărui subcorpus, înainte de producția propriu-zisă, apare un cod prin care este descris studentul. De pildă, *S001M-ENG-A1* se va citi: studentul 1, masculin, L1=engleză, grupa de vârstă A, textul 1 din proba de scriere. Amintim că am avut în vedere următoarele grupe de vârstă: A: 15-20 de ani, B: 21-25 de ani, C: 26-35 de ani, D: peste 35 de ani. Așadar, fiecare cod deci fiecare student, poate fi găsit o dată (o parte dintre studenții Erasmus) sau de două ori în corpus – o dată cu textul 1, a doua oară

cu textul 2. Avem două situații când un student poate fi găsit de 4 ori, fiind vorba de studenți care au repetat anul, producțiile acestora putând prezenta interes pentru o analiză comparativă.

#### 6.3.4. Conventii de transcriere

Corpusul de producții scrise ale vorbitorilor nonnativi de RLS, nivel A1, este un corpus primar, neadnotat, format din 56.185 de elemente și 5.625 de cuvinte diferite. Cu toate că nu e adnotat, în scopul păstrării confidențialității datelor cu caracter personal, am folosit o serie de convenții de transcriere. Amintim că studenții semnează un acord prin care își exprimă acordul/dezacordul cu privire la utilizarea producțiilor lor în cercetare, fapt ce determină și numărul de producții incluse în corpus.

Astfel, de fiecare dată când s-a făcut referire la un *nume străin*, care ar fi putut trimite spre autorul textului, am decis să înlăturăm acest nume și să îl înlocuim cu <Nume1/2/3/4....>. Există numeroase cazuri când, în cadrul aceluiași text, s-a repetat un nume. În această situație, am păstrat referința, pentru a păstra și coerența textului. În același mod am procedat și cu privire la orașul de poveniență sau la numele străzii din adresa subiecților. Acestea vor fi găsite în text astfel: <oraș/oraș1/2...>, <nume stradă> etc. Și în aces caz, dacă s-a repetat orice informație pusă inițial în <>, aceasta va apărea tot așa.

Pentru cazurile în care un cuvânt sau mai multe nu au fost inteligibile în momentul trascrierii, lipsa acestuia/acestora va fi semnalată în corpus astfel: <un cuvânt/două cuvinte>.

La localități, orașe, în general, am ales să păstrăm numele orașelor mari sau ale acelora care, din experiența de la grupele de studenți străini, știm că apar în discuții, prin urmare considerăm că sunt important de cunoscut. Să nu uităm că acest corpus poate fi folosit ca atare la cursuri, deci aceste referințe, să zicem, culturale, ar putea fi utile. La fel am procedat și în cazul țărilor sau al numelor. Am secretizat toate numele străine care ar fi putut trimite spre autorul textului și care ar fi putut viola clauza de confidențialitate, dar am păstrat toate numele românești, pentru a putea observa care sunt numele preferate sau reținute de studenți. De asemenea, nu am secretizat locurile din oraș, regiunile turistice, și nici numele străzilor Horea, Hasdeu si Piezișă, deoarece intenționăm să extragem o listă cu numele proprii, iar, în acest caz, considerăm că aceste informații din corpus sunt extrem de prețioase.

Întrucât, pentru analiza noastră, ne interesează exclusiv lexicul vehiculat de studenți, nu am avut în vedere probleme ce țin de segmentarea enunțurilor, ortografie și punctuație sau de gramatică, în general.

Tabelul 9 – Convenții de transcriere în corpusul de producții scrise ale vorbitorilor nonnativi, Nivelul AI

Convenție transcriere	Semnificație
<nume1,2></nume1,2>	s-a înlocuit numele original
<oraș></oraș>	s-a înlocuit orașul
< stradă>	s-a înlocuit numele străzii
<un cuvânt=""></un>	lipsește un cuvânt

Așa cum se și recomandă în literatura de specialitate, în scopul obținerii unor rezultate relevante și ilustrative pentru nivelul A1, în general, este important să avem în vedere și producțiile orale de la acest nivel. Astfel, în analiza noastră, pe lângă corpusul de producții scrise ale vorbitorilor nonnativi, utilizăm și CORLS, precum și un minicorpus format din cerințele materialelor de predare și evaluare, folosite la AP și E, ambele corpusuri fiind prezentate în continuare.

## 6.4. CORLS - Corpus oral de limba română ca limbă străină

Si CORLS<sup>618</sup> este un corpus specializat, de limbă vorbită, de nivel A1, compus din producțiile orale, transcrise în română, ca L2, de la 172 de studenți din AP, din cadrul a trei sesiuni de examinare, în perioada 2014-2017, format din 70.000 de cuvinte. În ceea ce privește subiecții, putem afirma că avem de-a face cu un gruptintă similar, preponderent vorbitor de L1 arabă, mai ales că o parte dintre subiectii din corpusul de limbă scrisă sunt reprezentați și în corpusul oral. Precum în cazul corpusului de producții scrise, și producțiile orale au fost colectate în cadrul examenului de nivel A1, ceea ce face ca din punctul de vedere al dificultății sarcinilor, să avem și în acest caz garanția că acestea sunt caracteristice nivelului A1. Proba orală a examenului A1 este formată din trei secțiuni: un interviu cu examinatorul (așadar, interacțiune pe teme cu caracter personal: familie, oraș, țara de provenientă, mâncare, activităti zilnice și de timp liber etc.), un monolog cu suport vizual (un set de imagini și o cerință cu câteva puncte de sprijin, pe teme specifice nivelului A1, prezentate în *Tabelul 10*, și o *conversație* între doi candidați, pe o temă dată, cu o cerință succintă, cu puncte de sprijin, enumerate în același tabel. Este de remarcat faptul că avem de-a face cu un discurs spontan, cel mult semiplanificat, în

<sup>618</sup> Descris pe larg în Lavinia-Iunia Vasiu, op. cit., 2020, pp. 130-140.

cazul monologului, ceea ce constituie un avantaj pentru analiza noastră, relevândune elementele lexicale însușite și stăpânite, deoarece doar pe acestea le va putea folosi corect candidatul în producția sa orală. Un alt element ce vine în sprijinul nostru este reprezentat de faptul că toate cele trei părți ale probei orale au un timp prestabilit, ceea ce va determina ca producția fiecărui candidat să fie reprezentativă pentru cunoștințele sale lexicale la nivelul A1. Să ne amintim, de exemplu, experimentul "asociativ-verbal" (prezentat în capitolul 4), care presupunea ca, pentru stabilirea vocabularului fundamental necesar în predarea unei limbi, să li se solicite subiecților primele maximum 20 de cuvinte care le vin în minte cu referire la un termen enunțat, într-un timp prestabilit. În *Tabelul 10*, prezentăm structura CORLS<sup>619</sup>:

Tabelul 10 – Structura CORLS

Sarcina	Timpul de	Timpul pentru	Reprezentare în corpus
	pregătire	producere	
Interviu cu examinatorul	neplanificat	1 minut	172 de studenți
Monolog cu suport vizual,	semiplanificat (1	2 minute	• familie (16 prod.)
imagini	minut de pregătire		• profesii (19 prod.)
	a răspunsului)		descriere de persoane
			(9 prod.)
			• vreme (19 prod.)
			• oraș (10 prod.)
			• casă (16 prod.)
			<ul> <li>vacanță (16 prod.)</li> </ul>
			mijloace de transport
			(11 prod.)
			mâncare și băutură (17 prod.)
			activități zilnice (17
			prod.)
			• timp liber (13 prod.)
			• îmbrăcăminte (9 prod.)
Conversație	neplanificat	3 minute	• la restaurant (16 prod.)
			organizarea unei
			petreceri (9 prod.)
			la magazinul alimentar
			(15 prod.)
			<ul> <li>planificarea unei zile</li> </ul>
			libere (15 prod.)
			• la agenția de turism (10 prod.)
			planificarea unei
			vacanțe (11 prod.)
			la agenția imobiliară     (15 prod.)

<sup>619</sup> Lavinia-Iunia Vasiu, op. cit., 2020, pp. 134-135.

Din parcurgerea temelor din cele două corpusuri, putem observa că sferele lexicale principale sunt reprezentate relativ echilibrat.

## 6.5. Corpusul de cerinte

Pe lângă cele două corpusuri ale vorbitorilor nonnativi de RLS, de nivel A1, am elaborat un mic corpus format din cerințele din manualele și materialele folosite în procesele de predare, învățare și evaluare de la Departamentul de limbă, cultură și civilizație românească, Facultatea de Litere, Universitatea Babeș-Bolyai. În crearea acestui corpus, am folosit cerințele din manualul principal utilizat atât la grupele de AP, cât și la cele de Erasmus, *Manual de limba română ca limbă străină (RLS). A1-A2*<sup>620</sup>, dar și din *Româna ca limbă străină (RLS). Caiete didactice. A1+*<sup>621</sup>, *Exerciții audio: Româna ca limbă străină (RLS): A1, A2, B1, B2, C1, C2*<sup>622</sup>, precum și din modelul de test de nivel A1, disponibil online, pe site-ul departamentului<sup>623</sup>.

Acest corpus a fost elaborat pentru a răspunde, la nevoie, criteriului necesității, adică pentru a ne asigura că, în inventarul lexical al RLS, nivel A1, includem și acele cuvinte care sunt esențiale în scrierea și înțelegerea cerințelor din cadrul cursurilor și al evaluărilor. Corpusul conține un număr de 7955 de cuvinte. Deși avem de-a face cu un corpus mic, considerăm că acesta este relevant pentru analiza noastră, întrucât cerințele nu comportă un grad de varietate atât de mare, mai ales în ceea ce privește evaluarea, unde cerințele din fiecare component al testului sunt aproape identice.

#### 6.6. Concluzii

În ciuda dimensiunii restrânse a corpusului de producții scrise ale vorbitorilor nonnativi de RLS, Nivel A1, considerăm că acest instrument reprezintă un pas înainte în asigurarea unor materiale elaborate prin analize empirice în domeniul RLS, recunoscându-i, cu toate acestea, limitele privind dimensiunea,

620 Elena Platon, Ioana Sonea, Dina Vîlcu, *Manual de limba română ca limbă străină (RLS). A1-A2*, Cluj-Napoca, Editura Casa Cărții de Știință, 2012.

<sup>621</sup> Platon Elena (coord.), *Româna ca limbă străină (RLS). Caiete didactice. A1*+, Cluj-Napoca, Editura Casa Cărții de Știință, 2012.

<sup>&</sup>lt;sup>622</sup> Elena Platon, Ioana Sonea, Dina Vîlcu, *Exerciții audio: Româna ca limbă străină (RLS): A1, A2, B1, B2, C1, C2*, Cluj-Napoca, Editura Efes, 2010.

<sup>623</sup> http://romaniandepartment.com/programe-de-studii/anul-pregatitor/modele-de-teste/.

varietatea producțiilor și a grupului-țintă. Totuși, cunoscându-se dificultatea cu care se colectează textele și se obține acceptul studenților de a li se folosi producțiile, suntem de părere că acest corpus poate fi considerat reprezentativ pentru grupul nostru țintă.

În secțiunea a II-a a lucrării prezentăm rezultatele analizei noastre, precum și inventarul lexical al RLS pentru nivelul A1.

# SECȚIUNEA A II-A – REZULTATELE CERCETĂRII

## CAPITOLUL VII – REZULTATELE CERCETĂRII

## 7.1. Metodologie

Pentru a obtine rezultatele cercetării noastre, am parcurs mai multe etape. Inițial, am efectuat o analiză, în principal, cantitativă a listelor obținute din corpusurile cu care lucrăm, după care am analizat, din perspectivă calitativă, lista rezultată după prima etapă în scopul obținerii unei forme finale a inventarului lexical al RLS pentru nivelul A1. În această a doua fază, am analizat concordantele pentru a vedea dacă elementele lexicale obtinute în prima etapă sunt adecvat si corect utilizate în context. Având în vedere că unul dintre principiile de elaborare a listelor de cuvinte este principiul replicabilității rezultatelor, încercăm să detaliem cât se poate de mult toate etapele parcurse. Un alt principiu se referă la cuvintele polisemantice si la omonime, omografe, omofone. Am luat în considerare si acest principiu, iar în etapa de analiză manuală am verificat și am identificat în corpus situațiile problematice. De asemenea, am încercat să evidențiem, în primul rând, ce poate face și deci înțelege, din punct de vedere lexical, un vorbitor nonnativ de limba română la nivelul A1, așadar, să arătăm ce aspecte ar trebui să caracterizeze microlimba la nivelul A1, si, mai putin, ceea ce nu poate face un vorbitor în acest stadiu, deoarece analiza noastră este una care urmăreste volumul vocabularului, acuratețea nefiind obiectivul nostru la acest nivel. Vom discuta totuși și despre ce anume nu este capabil să facă studentul, în special, pentru a propune alte elemente lexicale sau chiar transferul cuvântului la un nivel superior.

În acest scop, este important să decidem în ce moment considerăm că este însușit un anumit cuvânt. Mai mult decât atât, este esențial să avem în minte că, la acest nivel, nu putem vorbi totuși despre un control al vocabularului (vezi *supra*, cap. 1). În analiza noastră, folosim accepțiunea lui Nation privind cunoașterea lexicală (prezentată în cap. 5) și considerăm însușit un cuvânt (și deci inclus în inventarul lexical al RLS, nivel A1) dacă acesta este însușit atât din punctul de vedere al formei, cât și al conținutului, fiind utilizat corect în context. Având în vedere că ne referim la nivelul A1, am considerat însușit un cuvânt și dacă se dovedește a fi folosit adecvat în context, însă este scris greșit (de exemplu, apare literă mare în loc de literă mică, are inversate unele litere sau sunt substituite unele litere în cadrul cuvântului), *atâta vreme cât această eroare nu împiedică transmiterea mesajului*. Totuși, pentru astfel de situații, am decis introducerea în lista de frecvențe a unei coloane intitulate *alte* 

forme, unde vor fi inventariate diferitele forme corecte și greșite sub care apare un anumit cuvânt în corpus. Astfel, utilizatorul inventarului, dispunând de această informație, va putea acorda o mai mare atenție cuvintelor problematice, de exemplu. Astfel, dacă în analiza manuală a concordanțelor, am remarcat o utilizare corectă în cel puțin 50% dintre cazuri, considerăm elementul respectiv de nivel A1 și deci inclus în inventarul lexical al RLS.

Pentru o mai bună ilustrare a abilităților de scriere ale unui student de L2, așadar și pentru a avea o viziune cât mai clară asupra corpusului de producții scrise, am inclus un extras din grila de evaluare pentru producerea scrisă din *Companion Volume*. În *Tabelul 11* sunt marcate principalele trăsături ale L2 la nivelul A1. Subliniem că, la acest nivel, ne așteptăm la un **repertoriu minimal de cuvinte** și de **expresii memorate**, la **conectori puțini** și **foarte simpli** și la un **control extrem de redus** asupra cuvintelor și a expresiilor vehiculate, care sunt, mai degrabă, **memorate**. Menționăm că prin termenul de *conector* înțelegem elementul care asigură coeziunea și coerența textului, incluzând în această categorie atât conectorii sintactici, cât și pe cei pragmatici<sup>624</sup>.

Tabelul 11 - Grila de evaluare pentru producerea scrisă<sup>625</sup>

	A1					
Cunoștințe	Poate scrie expresii izolate sau enunțuri simple. Textele mai lungi					
generale	prezintă probleme de coerență ce fac ca textul să fie greu sau chiar					
	<u>imposibil</u> de <u>descifrat</u> .					
Volum	Are un <b>repertoriu minimal de <u>cuvinte și expresii simple</u></b> în legătură cu					
	detalii personale și situații concrete particulare.					
Coerență	Poate lega cuvinte sau grupuri de cuvinte cu conectori foarte simpli,					
	precum <i>și</i> sau <i>după</i> , <i>apoi</i> .					
Acuratețe	Arată un control extrem de limitat asupra unor puține structuri					
	gramaticale și <i>propoziții model</i> , dintr-un <i>repertoriu memorat</i> .					
	Erorile pot împiedica transmiterea mesajului. (sublinierea autorilor)					
Descriere	Poate <b>scrie</b> propoziții simple și enunțuri despre sine sau despre persoane					
	imaginare, unde locuiesc, ce fac etc.					
Argumentare	Nu există descriere nici la A1, nici la A2.					

<sup>&</sup>lt;sup>624</sup> Pentru diferența dintre conector sintactic și conector pragmatic, vezi *Gramatica de bază a limbii române (GBLR)*, pp. 7, 347, 658, 651-654.

<sup>625</sup> Companion Volume, pp. 173-174.

#### 7.2. Analiza cantitativă

În analiza noastră cantitativă folosim patru liste, de dimensiuni diferite, elaborate pe baza corpusurilor prezentate în partea 1 a lucrării.

Astfel, pornind de la corpusul de producții scrise ale vorbitorilor nonnativi, precum și de la CORLS, am obținut *o listă de frecvențe pentru limba scrisă*<sup>626</sup>, una pentru cea *orală*<sup>627</sup>, precum și o a treia *listă, generală*<sup>628</sup>. În această listă, au fost inventariate toate cuvintele cu forme distincte. Având în vedere dimensiunea redusă a corpusurilor, am elaborat aceste liste respectând două dintre cele trei principii statistice de creare a inventarelor, frecvența și răspândirea<sup>629</sup>, calculate electronic, după cum este prezentat în Capitolul 6. Cea de-a patra listă, folosită pentru a satisface criteriul necesității, este bazată pe un corpus foarte mic, format din cerințele din materialele RLS<sup>630</sup>.

Pentru cele trei liste bazate pe corpusurile nonnativilor, frecventa a fost calculată în două etape. În primă fază, am calculat FA, adică numărul de ocurențe ale unui cuvânt în corpus. Întrucât am analizat si corpusul CORLS, am considerat "cuyânt" fiecare segment ce este delimitat, în stânga și în dreapta, de un caracter special, adică: spatiu, virgulă, punct și virgulă, puncte de suspensie, paranteză etc. În această etapă, au fost socotite drept cuvinte și acele elemente numite în limba engleză fillers, adică modurile de umplere a pauzelor pline<sup>631</sup>, de tipul ăăă, eee, aaa<sup>632</sup>, specifice discursului oral si care, deloc surprinzător, ocupă un rang mare în lista generală: 2, 27, respectiv, 43. În cea de-a doua etapă, am calculat FR, stabilind baza de normalizare 100, întrucât vorbim despre un corpus restrâns ca dimensiune. Reamintim că frecventa se raportează la numărul de cuvinte din corpus, în cazul nostru, la suma elementelor din ambele corpusuri ale nonnativilor, așadar, calculele privind FA și FR sunt raportate la numărul **126.185 de cuvinte/elemente**. În graficul prezentat în continuare putem observa tendinta de descrestere a FA în lista generală de frecvente. Remarcăm o FA foarte ridicată pentru primele cuvinte, după care apare o descreștere abruptă, urmată de o oarecare constanță, într-un interval de 100 de valori, situație ce pare a fi în acord cu situația altor limbi.

*-* - -

<sup>626 5568</sup> de cuvinte indexate.

<sup>627 5372</sup> de cuvinte indexate.

<sup>628 9166</sup> de cuvinte indexate.

<sup>&</sup>lt;sup>629</sup> Amintim că în procesul de elaborare a listelor de cuvinte nu e nevoie să se folosească toate criteriile.
<sup>630</sup> Cu doar 1437 de cuvinte.

ea dour 1437 de eavinie.

631 Pentru mai multe detalii legate de modurile de umplere a pauzelor de către vorbitorii nonnativi de nivel A1, vezi Lavinia- Iunia Vasiu, *op. cit*, pp. 115, 149-150, 152, 153, 165.

<sup>632</sup> Sau alte variante ale acestora.

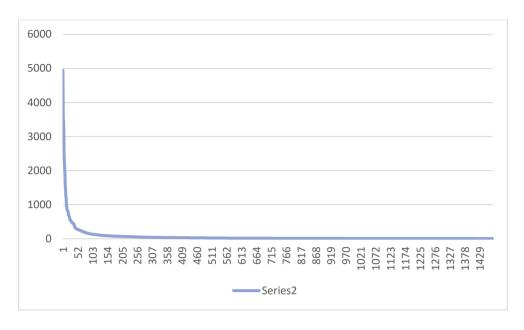


Figura 9 - FA în corpusurile vorbitorilor nonnativi de română ca L2

În ceea ce privește **răspândirea**, am calculat mai întâi răspândirea, în general, denumită R1, în liste, după care am optat și pentru calcularea R2-R4. Înainte de a explica demersul nostru, amintim că, spre deosebire de frecvență, răspândirea o raportăm la producție, nu la cuvânt. Dacă, în cazul corpusului de producții scrise, o producție presupune un text, în cazul celui oral, am considerat fiecare probă a examenului oral drept o producție. Amintim că, în toate cazurile, vorbim despre o producție axată asupra unei teme, exceptând producțiile din partea de conversație cu examinatorul, de la proba orală, care au mai degrabă un caracter general. De aceea, în final, am obținut un număr de **1196 de producții**. Astfel, vom citi R1 prin raportare la această valoare.

Pentru exemplificare, luăm primele 10 cuvinte, ca rang, din lista generală de frecvente, versiunea 1.

Tabelul 12 - Primele 10 cuvinte din lista de frecvențe, versiunea 1

Rang	Cuvânt	FA	FR	R1	R2	R3	R4
1	și	5244	438	1039	819	622	447
2	este	3974	332	1148	770	586	429
3	la	3161	264	903	676	485	332
4	în	2982	249	1195	776	439	248
5	eu	2591	217	764	534	390	279

6	de	2083	174	844	520	297	172
7	cu	1836	154	801	469	226	111
8	0	1497	125	623	324	178	113
9	un	1429	119	546	307	200	130
10	pentru	1257	105	590	304	158	78

Constatăm, asadar, că, în această etapă, cel mai frecvent cuvânt în corpusurile vorbitorilor nonnativi<sup>633</sup>, deci și în discursul acestora, este conjuncția și, cu FA de 5244 de ocurente și cu R1 de 1039, adică regăsim acest cuvânt în 1039 de texte cel putin o dată. Dacă facem comparatia cu Dictionarul de frecventă, care caracterizează discursul vorbitorilor nativi, cel mai frecvent element este prepoziția de (care în lista noastră ocupă rangul 6), însă, și în acest material, și are rangul 2. Mai mult, verificând și în lista de oral, și în cea de scris, constatăm că și ocupă tot primul rang. Rezultatul acesta nu ne surprinde mai ales că, la început, studenții întâmpină dificultăți în a-și structura discursul, iar și, de multe ori, cu toate că nu este necesar, stă în locul virgulei în producțiile lor (*Imaginea unu eu văd trei persoane*, un sofer, un sofer si un băiat si o femeie) sau chiar și a punctului (Du pă la munte suntem merge în club și noi avem encontran multe fete, și noi am dansat cu ele și după la club, noi suntem merge în cameră.), conducând la o suprautilizare a acestuia. Totuși, în general, cuvântul pare a fi bine însușit și folosit, cu foarte mici excepții (este inalt și slab, Acolo mănănc pasta si salata întotdeauna, este un frigider alb Şii un dulap galben). În ceea ce priveste răspândirea, locul întâi e ocupat de verbul a fi, cu R1=1850, urmat de prepoziția în (1195), de conjuncția și (1039) și de verbul a avea (1008).

Prin R2-R4 am dorit să aflăm în câte texte apare un anumit cuvânt de două ori, de trei ori, respectiv, de patru ori. După cum am arătat în prima parte a lucrării noastre, nu este suficient ca un cuvânt să fie doar frecvent, acesta trebuie să fie și răspândit, pentru a se evita situația melcului de mare, adică o concentrare a unui anumit cuvânt în cadrul unui text, iar, în altele, absența lui. Așadar, în opinia noastră, pentru a fi considerat răspândit, acesta ar trebui să aibă o bună reprezentare în R1, R2, astfel încât să se evite o suprafolosire a respectivului element, având în vedere și dimensiunea redusă a textelor din corpus.

Întrucât obiectivul nostru este acela de a obține un inventar lexical al RLS, general, cât mai relevant și cât mai ilustrativ pentru nivelul A1, am decis să excludem anumite cuvinte care, din punctul nostru de vedere, sunt specifice doar unui tip de

<sup>633</sup> Desigur, raportarea se va face în permanență la grupul nostru de subiecți, care, din punctul nostru de vedere, reprezintă grupul-țintă din AP și E de la Universitatea Babeș-Bolyai.

discurs (oral sau scris). Prin urmare, în analiza noastră, vizăm atât producerea, cât și receptarea, și discursul oral, dar și cel scris. De asemenea, luăm în considerare atât procesul de predare-învățare, cât și pe cel de evaluare. Pe lângă aceste elemente, vom exclude și substantivele proprii, care vor fi integrate într-o propunere de listă specială de nume proprii, dar și abrevierile, și acronimele.

În analiza noastră nu am avut în vedere literele alfabetului, pe care le considerăm din start necesare pentru a începe studiul limbii române ca limbă străină. Astfel, vom exclude din corpus și eventualele referiri la aceste litere. Un exemplu în acest sens este A1, utilizat nu pentru a face trimitere la nivelul A1, ci la unul dintre căminele din campusul studențesc. Cifrele, pe de altă parte, fac parte din categoria *sferelor lexicale* și vor fi tratate mai jos.

Din lista generală se vor exclude și eventualele scăpări rămase după analiza electronică a corpusurilor, ne referim, în special, la semnele folosite în transcrierea producțiilor orale, dar și la cuvinte preluate ca atare din alte limbi, mai ales din limba engleză: *yes*, *no*, *end* etc.

Cu toate că se observă o degradare a scrierii, odată cu scăderea frecventei, având în vedere că avem de-a face cu nivelul A1 și că dimensiunea corpusului nu este una foarte mare, am decis să excludem, într-o primă etapă, doar acele cuvinte care apar o singură dată în text. Acestea sunt, în general, cuvinte scrise greșit sau irelevante pentru nivelul A1, având în vedere că apar doar o dată, neputând influenta în vreun fel includerea sau excluderea anumitor cuvinte în/din inventar. Întrucât există și posibilitatea ca unele cuvinte să fi apărut și înainte în listă, pe o poziție superioară, scrise într-o altă formă, am decis să nu reducem, pentru moment, lista pe atât de drastic pe cât ne-ar indica-o frecvența. Am putut observa, că în funcție de modul în care au fost scrise unele cuvinte (cu sau fără diacritice, literă mare/mică, litere amestecate sau erori de scriere), unele elemente apar de mai multe ori în listă, pe poziții diferite, chiar și cele cu frecvență ridicată. Așadar, primul pas l-a constituit excluderea cuvintelor cu o singură ocurență, ceea ce a condus la obținerea unei liste generale inițiale de 3.709 cuvinte, care considerăm că ne asigură suficiente resurse pentru a identifica elementele lexicale specifice nivelului A1. Este important să ne amintim că, după calculele noastre, plecând de la situația altor limbi și a numărului de ore si de cuvinte care pot fi învătate pe oră, am decis să încercăm să ne situăm cu inventarul în intervalul 920-1428 de cuvinte.

Al doilea pas în analiză a fost excluderea din această listă de 3709 cuvinte a elementelor stabilite și amintite mai sus. Toate aceste cuvinte au fost inventariate în liste separate și vor fi prezentate mai jos.

După excluderea acestora, am observat că, în continuare, a rămas un număr mare de cuvinte, multe dintre acestea repetându-se sub diferite forme gresite ale unui cuvânt corect. În acest moment, am decis evaluarea manuală a listei, evidentiind formele scrise greșit, pe de-o parte, și pe cele care au probleme doar la scrierea cu diacritice, pe care, amintim din nou, nu le considerăm neapărat greșite, pe de altă parte. Mai mult, în situațiile în care au existat mai multe forme gresite, le-am căutat în corpusuri pentru a identifica dacă acele cuvinte au fost totusi folosite corect în context, iar, în funcție de concluzia noastră, am decis să le includem în listă, împreună cu forma scrisă corect, situație ce ridică valoarea frecvenței, sau, din contră, determină eliminarea din listă. Am păstrat în listă și variantele de scriere ale acestor cuvinte pentru ca utilizatorii listei<sup>634</sup> să aibă în vedere care sunt principalele erori de formă ce pot apărea în cazul acestor lexeme și să le preîntâmpine, pe cât posibil. Tot în acest moment, după analiza cu mâna și ochiul a listei, am decis ca toate cuvintele care au FA sub 4 să nu mai fie luate în calcul. Am luat această decizie, întrucât vorbim despre cuvinte a căror formă este foarte puternic afectată, ceea ce implică și o utilizare inadecvată, incapacitatea de a transmite sensul cuvântului din cauza alterării puternice a formei acestuia. Exemple: Biru, birul, (FA=3), Canabă, ucansă, caligii (FA=2). Prezentăm, în continuare, categoriile de cuvinte excluse din listă, oferind exemple sau alte propuneri de cuvinte, acolo unde se impune.

O altă etapă parcursă în drumul nostru spre obținerea inventarului lexical al RLS, pentru nivelul A1, este reprezentată de parcurgerea manuală a listei cu scopul de a verifica și de a confirma că nu s-au strecurat erori după analizele precedente. Astfel, am urmărit să nu mai existe variante ale aceluiași cuvânt la mai multe ranguri, apoi, am verificat dacă toți membrii unei paradigme gramaticale sunt luați împreună, și, nu în ultimul rând, am fost atenți să nu existe confuzii din cauza cuvintelor polisemantice. După parcurgerea acestei etape, am întregit sferele lexicale, așa cum ne-am propus inițial.

În continuare, analizăm cea de-a doua versiune a listei generale, evidențiind situațiile care ridică unele probleme, precum și pe acelea care necesită completarea sferelor lexicale. Întrucât vorbim despre un număr mare de elemente lexicale, considerăm nu este oportun să discutăm fiecare cuvânt, oferind exemple privind contextele de utilizare în cazul fiecăruia. În plus, având în vedere că inventarele lexicale nu sunt dicționare, nu oferim contexte sau definiții ale elementelor lexicale incluse în aceste liste decât în cazurile în care situația o impune. Totuși, dacă se dorește, se poate oricând verifica modul de utilizare al unui element lexical sau se

634 Cadrele didactice, în special.

pot evidenția contextele de utilizare ale acestuia prin consultarea corpusului de producții scrise ale vorbitorilor nonnativi.

În scopul obținerii inventarului lexical al RLS, în acest moment al analizei, am stabilit gruparea cuvintelor/formelor care aparțin aceleiași paradigme și care, în general, sunt predate împreună la clasă sub prima formă care apare în inventar. Este cazul formelor unui verb, la un anumit mod și timp, la diferite persoane, dar și al substantivelor și al adjectivelor. De exemplu, în cazul verbului *a fi*, cea mai frecventă formă este cea de indicativ, prezent, persoana a III-a, *este*. Astfel, la FA a acesteia am adăugat și valorile de la ceilalți membri ai paradigmei, *sunt, ești, suntem, sunteți*, prezentând, astfel, toate formele, așa cum sunt și predate și învățate, de altfel. Dacă, în cazul celorlalte categorii gramaticale, am optat pentru gruparea tuturor variantelor găsite în corpus sub aceeași formă de bază, la pronume, am decis să nu procedăm la fel, deoarece am fi crescut artificial frecvența acestora. De asemenea, în cazul verbului, am păstrat ca ocurență separată forma de participiu, întrucât aceasta presupune alte cunoștințe decât în cazul formelor de indicativ, de prezent sau de infinitiv.

Așa cum am observat din consultarea literaturii de specialitate, cele mai frecvente cuvinte sunt cele funcționale, situație ce se regăsește și în cazul nostru. După cum arătam și mai sus, conjuncția și reprezintă cel mai utilizat element lexical în corpusurile vorbitorilor nonnativi, urmată de verbul a fi, mai exact, de forma de indicativ, prezent, persoana a III-a, este, cu o FA de 3974. Rangul 3 este ocupat de prepoziția la, cu o FA de 3161, folosită în contexte precum: Sunt la mare, Maine mergem la muzeu, eu mănânc la cantină, În fiecare zi eu ma trezesc la opt ora, eu am curs de la ora 9 până la ora 13. Prepozitia în se află pe rangul al 4-lea si are FA de 2982 de ocurențe, găsindu-se în ambele variante scrisă și in (In fiecare dimineată eu mânânc in micul dejun lapte sau ceai si biscuit), iar, în corpusul oral, si en (o familie pilmba en mol, en la fotografia patru ăăă este in bucatari). Eu sunt în România de 7 luni, În weekend eu plec cu prietenii mei la mall, în fiecare zi eu ma trezesc la ora 07:00 dimineața, jucăm footbal în parc zilnic, merg în oraș, în pauză vorbesc cu prieteni mei, Sunt în restaurant acum. La rangul 5 este pronumele personal eu, cu FA de 2591, rezultat firesc, în opinia noastră, având în vedere că, la nivelul A1, raportarea se face în primul rând la sine și la situații curente. Ceilalți membri ai paradigmei au următoarele frecvențe: tu (FA=180), el (FA=475), ea (FA=482), noi (FA=279), voi (FA=70), ei (FA=279), ele (FA=25).

Următoarele în rang sunt prepoziția de, cu FA = 2083, utilizată în contexte precum: Facultate de litere, eu am curs de la ora 9 până la ora 13, Eu sunt din Romania de două luni., Mi-e dor de tine., Eu am nevoie de o ceai!, și conjuncția cu,

FA = 1836: merg la cinema cu prietenii mei, Îmi place carne cu cartof foarte mult., Eu merg la scoală cu autobuzul., este un taximetrist, ee, care vorbește cu o femeie.

La rangurile 8 și 9 se găsesc articolele nedefinite de singular, o și un, cu FA = 1497, respectiv, 1429, în vreme ce articolul nedefinit pentru plural, niște, are FA de 48 (ascultam muzica Rock și ne uităm la niște fîlme). Pe locul 10 se găsește prepoziția pentru, cu FA = 1257, scrisă și pentru și pântru. În corpusul scris, apar 669 de ocurențe ale prepoziției pentru, dintre care 402 sunt, de fapt, pentru că. Așadar, contextele de apariție ale lui pentru sunt: În fiecare zi eu merg la Facultatea de Litere pentru că am curs de limbă Română (cauză), Apoi mă culc un pic și fac tema și exerciții pentru mâine., lucrez la curs acasă și pentru cina, iau spaghete cu pui sau supă., În România eu studiez limba romana pentru un am. (durată), eu mă duc la mall cu prietenii mei pentru cumpăraturi (scop).

Verbul *a merge* pare a fi unul extrem de prolific la nivelul A1, găsindu-se în 1139 de situații, fiind scris sub următoarele forme în corpus: merge, mergem, mergeți, mergi, merga, merj, merjem, merji. Contextele în care acest verb poate fi întâlnit la acest nivel sunt: apoi mergem la sala de sport., Vrem să mergem la facultatea de litere., că Nu îmi place merg Pe jos., Astăz vom merge să măncâm pizza., Dupe ce dorm, mergem cu familie mai in centru sau vedem un film în cinema. O situație similară întâlnim și în cazul verbului a-i plăcea, care are, în această versiune a listei, FA = 1001. Dacă, în corpusul scris, acesta se regăsește corect scris (Mie Îmi place plăcintă., Îmi place să schiez in iarna., Nu îmi place orașul., Nu imi plac animalele în camera.), în cel oral, apare și cu forma plașe (imi plașe vara, Nu eu îm plașe mare.), cu puține ocurențe, la unii vorbitori de spaniolă și de franceză. Ar putea părea prea frecvent acest verb, însă, totuși, dacă verificăm și în Dicționarul de frecvență, observăm că ocupă rangul 433, așadar, constituie un element lexical important și din perspectiva vorbitorului nativ. Din punctul nostru de vedere, rezultatul este unul relevant pentru situația noastră particulară, întrucât, în lipsa altor verbe, poate, uneori, mai potrivite, la nivelul A1, a-i plăcea este des utilizat. Întradevăr, FA poate fi ușor influențată și de temele de scriere și de vorbire în care li se solicită candidatilor să numească ce le place și ce nu le place în casa/camera lor, de pildă, sau în imaginile de la proba de monolog.

Adverbul *nu* este mai frecvent decât *da*, cu 950 față de 807 ocurențe: *Dar Nu îmi plac fotolii pentru că sunt murdări, dar nu am bani, Nu știu foarte multe din din Constanța*. Am constatat folosirea lui *nu* pentru a indica antonimul care de cele mai multe ori presupune un nivel mai avansat în limbă: *nu este etil* (=inutil), *Nu e confortabil* (inconfortabil), *Asta nu e linișteta* (= zonă zgomotoasă). În opinia noastră, e firesc ca *nu* să fie mult mai frecvent decât *da*, fiindcă acesta e folosit și la

forma de negativ a verbelor. Da, pe de altă parte, are drept prim context exprimarea afirmației, cu o prezență covârșitoare în corpusul oral, forma da apare și în locul prepoziției de (cu merg la facultata in fiecare zi da ora noua și intorc acasa la ora uno). Considerăm, așadar, că FA nu este cea relevantă. Cu toate acestea, necesitatea acestui adverb nu poate fi contestată. Prin urmare, considerăm că da ar trebui să rămână inclus în inventar.

Forma de singular, feminin, a adjectivului pronominal posesiv, *mea*, este cea mai utilizată, având o FA de 839 de apariții, fiind regăsită în corpus și sub forma *mia* (*in seara mananc supa cu legume tot in acasă mia.*, *mama l- mia*).

Conjuncția *că* ocupă și ea un loc important, 22, însă sunt luate în calcul la obținerea FA următoarele situații: membru în locuțiunea conjuncțională *pentru că*, dar și mijloc de introducere a subordonatelor: *cred că*, *sper că*, *îmi place că*, *să spuneți că*. Am constatat că, în corpusul oral, aceasta apare și sub forma lui *che: Îmi place Spania, pentru che locuiesc in Franța*, *este normal que*, *ăăă*, *che femei juacă futbol/football*.

FA a substantivului *cameră* credem că a fost influențată de una dintre temele de scriere, *Descrierea camerei*, apărând în listă cu 670 de ocurențe și un R1 de 333. Să ne amintim că avem 153 de producții pe tema *casa*, în corpusul scris, și 31 de producții, în cel oral, dacă luăm în calcul și tema *La agenția imobiliară*. Ne așteptăm, așadar, la obținerea unei FA mai ridicate, care nu reprezintă o problemă decât, poate, în ceea ce privește rangul acestui cuvânt. Totuși, încă o dată, subliniem necesitatea regăsirii acestuia în inventarul lexical al RLS.

La rangul 31, apare cuvântul *mare*, cu FA de 543 de cuvinte, cu două lexeme (*un oraș mare* în Germânia, El are douazeci și doi de ani și este student în un oraș mare de Franța., Îi place la mare și la soare în vară.). Și la rangurile 638 și 639 găsim cuvintele mar (FA=8) și mari (FA=8), care presupun ambele lexeme. Întrucât, la acest nivel, este vorba atât de substantiv, cât și de adjectiv, am investigat exemplele din corpusuri pentru a observa cum sunt utilizate cele două lexeme. Am constatat că substantivul mare e însoțit aproape întotdeauna de prepoziția la: apoi mergem la mare, înregistrând o FA de 163 din cele 543 de apariții. În listă este înregistrat și substantivul marea, cu FA=19, dar și mară, cu FA=12, cu același sens, la mare. Astfel, vom avea în listă două intrări: mare (adj), cu FA=388, și mare (subst.), cu FA=202.

Cuvântul *bine* are 489 de ocurențe în lista generală. Totuși, am observat că o mare parte dintre acestea nu reprezintă utilizări corecte ale adverbului *bine*, ci sunt forme greșite ale adjectivului *bun* sau ale adverbului *mult*. Astfel, în corpusul scris apare confuzia *bun-bine*: *eu gatist e foarte bine mâncăre*, *Mancarea este bine și* 

sănătoasă., Dacă vremea e bine (soare și un cer senin), mergem la Parcul Central, <Nume1> este foarte bine persoane, ea foarte bine cu oamenii. Pe de altă parte, în corpusul oral, întâlnim confuzia mult-bine: "mie imi place foarte bine, mie imi place foarte bine la cabana pentru ca eu văd, eeee, toate, tot, toți copac, copaciii, toți copaciii, em, în padure, în padure, eee, un lac, eeee (4s) animales, animale". Totuși, formele corecte sunt mai multe decât cele greșite, ceea ce ne face să concluzionăm că aceste elemente lexicale sunt potrivite pentru nivelul A1, însă necesită o atenție specială în cadrul cursurilor. Ca urmare a acestei situații, am decis înjumătățirea FA, pentru a putea ilustra mult mai corect rangul acestui adverb. Remarcăm că adjectivul bun (FA=309) este unul ce caracterizează nivelul A1, folosit, în general, adecvat în context, însă cu probleme, în special, la acord: Avem a zi bun cu prietenii mei; Tata mea este o persoană mai bun, este foarte bun preparator pește.

Elementul lexical drag caracterizează, mai cu seamă, discursul scris, regăsindu-se în corpus drept: substantivul dragul/draga, adjectivul drag, dar face parte și din locuțiunea adverbială cu drag. Astfel, îl identificăm în corpus cu forma dragă, 399 de ocurențe, precum și cu variantele draga, drăga, drăgă. Este folosit, în principal, în formula de început (*Dragă* <*Nume1*>!, *Draga Dragă* <*Nume1*>, *Salut*, dragă <Numel>, Dragă familia mea, Draga mea familie) și în cea de încheiere (de cele mai multe ori greșit!) de la textul cu tipar interacțional: Cu dragă drag, <Nume2>, Cu dragă <Nume2>, Cu draga, <Nume1>, Te pup, draga <Nume2>, Multumesc, drăgă, la revedere!. Lexemul drag, cu 277 de ocurente, apare, în special, în formula de final, cu drag, utilizat corect. Cu un număr mult mai mic de ocurențe, apare dragii (11 ocurențe), care este corect utilizat în sintagma posesivă dragii mei, dragi (8 ocurențe), folosit greșit, în locul formei articulate sau în situația: Dragi Mama și Tata!, precum și dragul, cu 7 ocurențe, în general folosit corect. Din exemplele identificate în corpus, sintagmele Cu drag, Dragă...., Dragi ..., Dragii mei par a fi memorate, însă, de multe ori, greșit, mai ales în ceea ce privește formula de încheiere. Din acest motiv, credem că nu e oportun să se introducă la nivelul A1 și substantivul, și adjectivul, dar și locuțiunea adverbială *cu drag*. În opinia noastră, aceste lexeme ar fi mai potrivite la A2, iar la A1, la scriere, s-ar putea opta pentru o formulare mai simplă (poate o adresare direct cu numele și, la final, utilizarea unei formule mult mai frecvent utilizată, precum te/vă pup), având în vedere că, la acest nivel, textele scrise sunt extrem de simple.

Substantivul comun *parc* pare a fi însușit corespunzător: *alerg într-un parc ăăă ai ei are, merg în parc, În fiecare zi eu alerg în parc.*. Îl regăsim în lista generală cu 218 ocurențe, dintre care 8 cu forma *barc*, în corpusul oral, la vorbitorii cu L1 arabă (*plimbare la barc.*, *batru persoane în barc mânâncați frute*). Dacă, în cazul

structurii *merg în parc*, aceasta este, în general, adecvat utilizată, în cazul combinării cu alte verbe, pare a exista o nesiguranță la selecția prepoziției adecvate (*în/la*) cu substantivul *parc*: *Apoi eu merg să plimba la parc cu prietenul meu.*, *după facultate mă duc cu prietenii mei la parc sau în oraș.*, *Eu ma vizitat la parc în centru.* 

Cuvântul *obicei*, ca parte a locuțiunii adverbiale *de obicei*, la acest nivel, are FA de 159 de ocurențe: *De obicei eu mă întâlnesc cu prietenii mei, noi ne plimbăm, noi mergem la o petrecere, De obicei el poartă hainele sportive, De obicei pregatesc micul dejun, pranzul și cină.*, La pranz mananc de obicei Pastele cu rosii sauce.

Cuvântul *fata* e folosit în corpus atât pentru substantivul *fată* (de cele mai multe ori), cât și pentru *față* (*mă spăl pe față*). Pe lângă aceste utilizări, am constatat că se încearcă, destul de des, folosirea locuțiunii prepoziționale *în fața*, de fiecare dată, greșit, având în vedere că este o structură gramaticală ce se introduce doar la finalul nivelului A2: *un balcon fața strada*, *În fața de pat*.

În ceea ce privește verbul *a (se) uita*, remarcăm faptul că forma sa participială, *uitat*, este utilizată în locul celei de indicativ prezent: *eu merg acasa și mă uitat la televizor*, *cinma uitat un film*.

Substantivul *paste* pare a fi unul problematic, din cauza asemănării cu limba engleză, *pasta* (*Eu știu să gătesc prajitura*, *legumele cu orez, pasta și mai mult feluri de mâncare.*, *are un Pasta patru fromage foarte bun.*). Astfel, am constat că, în corpus, se folosesc ambele forme (*paste*, *pasta*), aproximativ în aceeași proporție. Considerăm că acesta este un cuvânt adecvat pentru nivelul A1, însă credem că este important să se accentueze diferenta fată de forma din limba engleză.

U în cele două corpusuri este folosit în locul articolului nedefinit o (u rochie), dar și pentru a marca ezitarea, în corpusul oral: cumPĂrăm... u u... un calculator sau un telefon, însă u u îi place. Din acest motiv și având în vedere că articolul este suficient de bine reprezentat, am decis eliminarea lui u din listă, fără a adăuga frecvența în altă parte. Aceeași decizie am luat-o și în cazul lui me, care e utilizat atât în locul formei pronominale mă (me trezesc), cât și îmi (me place), dar și, în unele cazuri, în locul adverbului mai (ce me faci?). O situație similară o reprezintă și cazul lui mi, care desemnează atât pronumele meu, mie, cât și adverbul mai. În corpusul oral, am întâlnit forma populară io a pronumelui personal eu: Acum in Cluj io vin cu autobuzul. Și în această situație am luat decizia excluderii din listă a acestui cuvânt, din cauza apariției acestuia, mai cu seamă, în corpusul oral.

Particula *ar* apare cu un număr de 16 ocurențe, desemnând de cele mai multe ori verbul a avea (*El ar 170m*, *Ar câteodată nu are multă lumina*, *ei ar cumpărat multă lucruri*, *ește ar timp liber*), însă, în general, este folosită greșit, ceea ce ne-a determinat să nu luăm în considerare aceste forme. Conjuncția *ci* nu e folosită corect,

ceea ce e și normal la acest nivel, având în vedere că, după *Descrierea minimală*, se introduce abia la nivelul A2<sup>635</sup>. În corpusurile vorbitorilor nonnativi apare în locul conjuncțiilor și și că (Suc, eu prefer de mere ci portocale., Pentru desert eu pereferat supa de legume pentru ci este foarte santate.), a pronumelui ce (Din ci țară ești). În plus, am decis și excluderea lui ie din cauza incoerenței: îi ie place foarte mult să gatească.

La o ultimă parcurgere a listei, am decis excluderea unor elemente din cauza folosirii greșite de fiecare dată (posibilat, sinci, pudem, perencii, papa, oarma, mulce, li, cunde, ame, bau, îngheață în loc de înghețată, famie, ba), a utilizării doar în formulele memorate (al), a preluării unor cuvinte ca atare din alte limbi (soda, relax, is, rumana, market, foto, can, moto, es, esta). Cele mai multe dintre cuvintele excluse fac parte din categoria cu FR mai mică decât 1. După cum am văzut în prima parte a lucrării noastre, limita recomandată (cut-off) este tocmai la FR=1. Totuși, având în vedere că nu am dispus de un corpus prea extins, am observat că multe dintre cuvintele importante, din punctul nostru de vedere, s-ar găsi sub cut-off (de exemplu: august, decembrie, februarie, băutură, inginer, preț, anotimp, călătorie, caiet, cap etc.), de aceea am decis selectarea lor, în funcție de necesitate și ținând cont și de apartenența lor la categoria cuvintelor internaționale (de pildă: grup, inginer, acțiune, biscuit, chitară, computer, dezavantaj, direct, farmacie, internet, etc.).

#### 7.3. Substantivele proprii

Așa cum am subliniat în partea teoretică (vezi *supra*, 5.3.), necunoașterea numelor proprii poate împiedica înțelegerea. Din acest motiv, precum și datorită numărului mare de substantive proprii care apar în corpusul scris, propunem formarea unei/unor liste suplimentare, cu caracter deschis, în care să se găsească aceste elemente. Considerăm că, în funcție de grupul-țintă, precum și de centrul universitar în care va fi folosit inventarul lexical, substantivele proprii vehiculate vor fi altele, adaptate situației. Avem în vedere, în această analiză, următoarele categorii: continente, țări, orașe, nume de persoane, personalități (în corpusul scris, doar Angelina Jolie și Picasso) și locuri din Cluj. Pentru a reduce din variante și având în vedere că nu considerăm greșeală, în acest moment, scrierea cu literă mică în loc de

<sup>635</sup> Cu toate că în predare se introduce, cel puțin în Platon, Sonea, Vîlcu, *op. cit.*, 2012, în cadrul unității 8, de pildă. Menționăm că prin decizia noastră de a nu include acest element lexical la acest nivel (dar și pe altele, despre care am discutat deja), nu susținem că aceste cuvinte nu ar trebui predate studenților, ci că nu ar trebui să le presupunem însusite de către acestia, la nivelul A1.

mare, toate substantivele proprii au fost generate în listă cu literă mică pentru a putea fi reperate toate variantele de scriere ale unei forme. Noi le inventariem, aici, corect. În corpus, acestea rămân asa cum au fost scrise de către studenti.

Cele mai multe substantive proprii sunt din categoria țărilor, iar prima țară ca frecvență este *România* (FA=72), care apare în lista cu frecvențe cu mai multe variante: *romania* (49), *româniă* (FA=8), *romănia* (FA=5), *rumeinia* (FA=3), *rumania* (FA=2), *româniâ* (FA=2). Astfel, chiar și dacă ignorăm variantele greșite, cu excepția lipsei diacriticelor, tot avem o frecvență bună a acestui cuvânt, de 121 de ocurențe, ceea ce face să fie inclus în inventar. O altă țară cu multe variante de scriere este *Franța*: *franța* (FA=10), *franță* (FA=7), *frața* (FA=2) *franta* (FA=2). Exceptându-le pe acestea și scrierea în limba engleză a numelui *Turciei*, celelalte denumiri sunt corecte, ceea ce ne face să confirmăm necesitatea cunoașterii țărilor la nivelul A1.

În ceea ce privește continentele, nu se remarcă o frecvență ridicată. Totuși, considerăm că este important să fie întregită lista și să se cunoască continentele, însă aceasta va fi, desigur, alegerea fiecărui utilizator al inventarului lexical al RLS, Nivelul A1.

Nici numele orașelor nu sunt folosite foarte des. Totuși, observăm că orașele românești apar de multe ori scrise greșit. Astfel, primul, după cum e firesc în cazul nostru, apare *Cluj* (FA=296), *Cluj-Napoca* (FA=65), *Napoca* (17) și alte variante, precum: *clujul* (FA=12), *cluj-* (FA=2). Următorul oraș este *București*, scris *bucarești* (FA=11), *bucurești* (FA=6), dar și *Brașov* (scris și *Brașnov*) sau *Constanța* (scris și *cotanta*) sau *timișoara* (scris și *timișoira*), precum și *Alba Iulia* (cu 4 ocurențe). Din punctul nostru de vedere, familiarizarea străinilor cu numele românești ale orașelor din România, dar și din străinatate, e utilă, mai ales în situația studenților care învață româna ca L2 în mediu vorbitor și care se vor mai întâlni cu aceste denumiri și în alte contexte.

Din dorința familiarizării cât mai rapide și mai eficiente a vorbitorilor nonnativi cu orașul Cluj-Napoca, am izolat cuvintele care fac trimitere la oraș sau la alte locuri din (apropierea) oraș(ului) care pot prezenta interes. Un alt motiv al acestei selecții l-a constituit și identificarea locurilor care prezintă cel mai mare interes pentru studenți pentru a le avea în vedere cu viitoarele generații de studenți. Am remarcat că cel mai frecvent cuvânt e numele străzii *Hasdeu*, unde se găseste campusul studențesc (FA=35), numele Universitășii *Babeș-Bolyai*, scris cu mici erori de scriere, (FA=18), respectiv, *Iulius mall* (14).

După cum am explicat în descrierea corpusului de producții scrise, am optat pentru păstrarea numelor românești clasice, care nu ar putea trimite spre autorul

textului, în scopul identificării numelor cu care sunt mai familiarizați vorbitorii nonnativi. Observăm că cel mai frecvent nume românesc în corpusul nonnativilor este *Maria*, nume ce apare des și în materialele de predare și evaluare. Totuși, în ciuda insistențelor de la clasă, remarcăm lipsa numelor de persoane, de gen masculin, cu final –*a: Luca, Horea, Mircea*, si a celor feminine cu finalul în consoană: *Carmen*.

Tabelul 13 – Lista numelor proprii care se întâlnesc în corpusul de producții scrise ale vorbitorilor nonnativi de RLS

Cuvânt	FA	FR	R1	R2	R3	R4
Maria	18	0.015	15	3	0	0
Ana	8	0.007	6	1	1	0
Mihai	7	0.006	7	0	0	0
Alex	7	0.006	6	1	0	0
Andra	6	0.005	3	1	1	1
Andrei	4	0.003	4	0	0	0
Andreea	3	0.003	2	1	0	0
Alina	3	0.003	2	1	0	0
Alexandra	1	0.001	1	0	0	0
Elena	3	0.003	3	0	0	0
Ina	3	0.003	2	1	0	0
Daria	3	0.003	1	1	1	0
Ciprian	2	0.002	1	1	0	0
Gheorghe	2	0.002	2	0	0	0
Diana	2	0.002	2	0	0	0
Vasile	2	0.002	2	0	0	0
David	2	0.002	2	0	0	0
Robert	2	0.002	2	0	0	0

## 7.4. Acronime și abrevieri

În urma analizei au rezultat 7 acronime și abrevieri, izolate din lista generală și prezentate în tabelul de mai jos:

Tabelul 14 – Lista acronimelor și a abrevierilor care apar în corpusurile vorbitorilor nonnativi de RLS

C	Tr A	ED	D1	D.O	D2	D4
Cuvant	FA	l FK	KI	<b>R</b> 2	K3	K4

nr	6	0.005	6	0	0	0
cm	4	0.003	4	0	0	0
fspac	3	0.003	3	0	0	0
kg	3	0.003	3	0	0	0
umf	3	0.003	3	0	0	0
etc	2	0.002	2	0	0	0
ubb	2	0.002	2	0	0	0

Cu toate că nu au o frecvență ridicată, considerăm că unele dintre acestea sunt relevante pentru grupul nostru țintă (nr., cm., kg., etc.), alături de m, km și str., la fel de utile ca primele. Credem că sunt necesare aceste abrevieri, mai ales în contextul în care învățarea are loc în mediu vorbitor, iar studenții vor lua contact destul de repede cu aceste prescurtări.

Coroborând listele de mai sus, obținem, în *Anexa 1*, o propunere de listă de substantive proprii pentru româna ca L2, nivelul A1, în care am marcat cu gri elementele inserate de noi în scopul întregirii anumitor categorii. Amintim încă o dată caracterul pur orientativ (dar și specific grupului nostru țintă) al acestei liste, care poate fi modificată în funcție de grupul-tintă.

## 7.5. Numerele

Numeralul cardinal este foarte prezent în corpusurile vorbitorilor nonnativi de RLS, ceea ce presupune o bună reprezentare și în lista generală de frecvențe. Acesta este motivul pentru care am decis extragerea tuturor numerelor și formarea unei liste separate, care va fi reluată și în momentul discutării criteriului sferelor lexicale. La acest nivel, numerele apar scrise atât cu cifre, cât și cu litere. Pentru a asigura unitatea listei noastre, optăm pentru scrierea cu litere în inventar.

Tabelul 15 – Lista numerelor care apar în corpusurile vorbitorilor nonnativi de RLS

Cuvântul	Variante sub care apare în corpusuri	FA	FR	R1	R2	R3	R4
zero	0	162	453	80	41	21	11
unu, una	1, uno, une	255	440	177	62	14	2
doi, două	doua, dua, doa, dou, doi, doua	676	249	445	147	56	17
trei	3, trei, tre	334	438	208	67	29	17
patru	4, patra, patru,batru	245	164	160	52	22	7
cinci	5, cinci, cince	191	127	129	39	13	5

	T		1		1	1	1
șase	6	88	264	48	14	5	2
şapte	7	99	4.1	81	6	3	2
opt	8	128	154	108	18	2	0
nouă	9	62	5.2	61	1	0	0
nouă		40	95	34	5	1	0
zece	10, zeci	86	125	73	11	1	1
unsprezece	11	31	105	28	3	0	0
doisprezece	12	32	84	28	4	0	0
treisprezece	13	68	4.9	66	2	0	0
paisprezece	14	13	119	12	1	0	0
cincisprezece	15	13	59	13	0	0	0
şaisprezece	16	15	79	15	0	0	0
şaptesprezece		15	48	14	1	0	0
optsprezece	18	34	40	33	1	0	0
nouăsprezece	19	17	55	16	1	0	0
douăzeci	20, douăzece,douazeci	111	49	94	14	3	0
douăzeci și unu	21	12	78	12	0	0	0
douăzeci și doi	22	8	35	8	0	0	0
douăzeci și trei	23	15	51	15	0	0	0
douăzeci și patru	24	6	33	6	0	0	0
douăzeci și cinci	25	12	48	12	0	0	0
douăzeci și șase	26	4	56	4	0	0	0
treizeci	30	74	4.5	59	12	3	0
patruzeci		8	33	8	0	0	0
patruzeci și cinci	45	10	45	9	1	0	0
cincizeci	50	26	54	25	1	0	0
cincizeci și șase	56	4	31	4	0	0	0
şaizeci	60	11	51	9	1	1	0
şaptezeci		4	37	2	1	0	0
optzeci		4	29	3	1	0	0
nouăzeci	90	6	32	6	0	0	0
o sută	100, sute	24	46	16	3	1	0
mii	·	7	34	4	3	0	0
două mii							
cincisprezece	2015	6	46	6	0	0	0

## 7.6. Excluderi din lista de frecvențe

#### 7.6.1. Mărci ale discursului oral

Am exclus definitiv din listă două categorii de elemente care sunt specifice discursului oral, de tipul pauzelor pline sau al metodelor de marcare a ezitării, așanumitele *începuturi false*<sup>636</sup>. În general, aceste mărci ale oralului sunt specifice limbii materne a vorbitorului, țin de strategiile discursive ale acestuia, și nu trebuie învățate într-un cadru formal. Prezentăm, în continuare, o parte dintre elementele excluse, în ciuda frecvenței ridicate pe care au înregistrat-o, datorită ponderii mari pe care CORLS a avut-o în generarea listei:

Tabelul 16 – Elemente ce aparțin discursului oral, excluse din analiză

Cuvânt	FA	FR	R1	R2	R3	R4
ăăă	3564	2.98	292	270	244	231
eee	546	0.457	109	84	71	52
aaa	341	0.285	82	49	33	24
a	248	0.207	161	42	21	12
ăăăm	102	0.085	27	16	12	9
mmm	94	0.079	50	22	10	3
ăm	76	0.064	28	15	9	8
((râde))	73	0.061	37	17	7	4
iii	50	0.042	26	14	8	2
ah	29	0.024	24	4	1	0
hm	23	0.019	17	4	2	0
ăăăă	23	0.019	18	5	0	0
îîî	20	0.017	13	5	1	1
îhâm	18	0.015	11	4	2	1
ăă	17	0.014	14	3	0	0
ă	17	0.014	14	2	1	0
ee	16	0.013	13	3	0	0
uuu	15	0.013	12	3	0	0
h	11	0.009	10	1	0	0
000	10	0.008	9	1	0	0
emmm	10	0.008	5	2	2	1

<sup>&</sup>lt;sup>636</sup> Pentru mai multe detalii legate de începuturile false în cazul IL vorbitorilor nonnativi de nivel A1, vezi, Lavinia-Iunia Vasiu, *op. cit.*, 2020, cap. III.

175

După cum se poate observa din tabel, pe lângă aceste elemente, am identificat și unele ce țin de convențiile de transcriere (de exemplu:  $(r\hat{a}de)$ ) și care au fost omise la analiza primară, realizată cu ajutorul calculatorului. Toate aceste elemente reperate, precum și alte variante mai puțin frecvente ale mărcilor discursului oral (em, aha, aa, e, mda, pf, lol, hmm etc.) au fost excluse din aceleași motive: fiind specifice oralului, ar influența mult prea tare inventarul lexical, putând chiar ocupa locul unui cuvânt mult mai productiv în comunicare, în general.

Pa-pa, fa-, în-în do-, s- ("În imaginea doi, ăăă, s-, este fameie cu..., Da, s-, (4s) soar. Ăăă (5s) pe s-... pe scaune" false ce au fost identificate în corpus. În consecință, de fiecare dată când au fost reperate în text, au fost eliminate din lista generală.

## 7.6.2. Transferuri din alte limbi

Cu toate că fenomenul transferului lexical este extrem de prezent în corpus, situație normală și de așteptat la acest nivel, am marcat și am exclus, în această etapă, doar acele cuvinte care sunt transferate întocmai dintr-o altă limbă, în special, din engleză. Am optat pentru această selecție întrucât celelalte transferuri, mai puțin evidente, ar putea fi utile în etapa următoare a analizei pe lista generală. Așadar, am exclus cuvinte precum:

Cuvânt	FA	FR	R1	R2	R3	R4
kilo	14	0.012	5	3	2	2
and	12	0.01	10	2	0	0
sorry	9	0.008	7	1	1	0
to	9	0.008	7	2	0	0
end	7	0.006	6	1	0	0
be	6	0.005	4	2	0	0
it	6	0.005	6	0	0	0
		0.005	5	1	0	0
yes	6			1	_	
oui	5	0.004	4	1	0	0
action	4	0.003	4	0	0	0
my	4	0.003	4	0	0	0

Tabelul 17 – Transferuri directe dintr-o limbă străină

<sup>637</sup> CORLS, S22, S25, S168.

party	4	0.003	3	1	0	0
know	4	0.003	4	0	0	0
don't	4	0.003	4	0	0	0
what	4	0.003	4	0	0	0
footbal	4	0.003	4	0	0	0
game	4	0.003	4	0	0	0
this	4	0.003	3	1	0	0
again	4	0.003	4	0	0	0
on	4	0.003	2	1	1	0

Precizăm că lista cuvintelor excluse este mult mai extinsă, noi optând pentru evidențierea cuvintelor cu o frecvență mai mare. Remarcăm preponderența transferurilor dinspre limba engleză, existând puține reprezentări din alte limbi: oui, palov. Dacă ne raportăm din nou la variabilele individuale, mai precis, la limbile străine cunoscute, se explică de ce sunt majoritare transferurile din limba engleză. Mai mult, limba engleză este și limba de contact, limba administrativă, la început, în cadrul comunicării cu studenții străini.

## 7.6.3. Caractere speciale sau litere

După cum spuneam mai sus, nu ne interesează includerea literelor alfabetului în inventarul lexical al RLS, nivel A1. Din acest motiv am eliminat din lista generală toate literele care au apărut izolat. Pe lângă acestea, am omis și alte caractere speciale care s-au găsit destul de frecvent. Exemple în acest sens se găsesc în tabelul de mai jos:

Cuvânt FA FR R1 R2 R3 R4 a1 20 0.017 20 0 0 0 0 0.009 10 1 0 11 0 0.009 2 0 11 9 9 0.008 0 0 0 6 6 0.005 0 0 0 p  $\odot$ 0.003 0 0

Tabelul 18 – Transferuri directe dintr-o limbă străină

Se remarcă pe prima poziție în tabel numărul unuia dintre căminele studențești în care sunt cazați, în general, studenții străini. Totuși, nu considerăm că acesta reprezintă un motiv suficient de întemeiat pentru a include în inventarul de

A1 aceste referințe. Mai mult, dacă am include un element, ar trebui completat grupul de cuvinte cu toate celelalte numere ale căminelor, ceea ce nu credem că ar fi oportun.

#### 7.7. Analiza calitativă

După faza cantitativă, în care am redus semnificativ lista, așa cum am prezentat deja, în etapa calitativă, am analizat concordanțele pentru a identifica expresiile sau grupurile de cuvinte folosite recurent de către studenți în **producțiile lor scrise**, evidențiind formulările învățate "după ureche", expresiile, colocațiile sau alte grupuri de cuvinte memorate și utilizate frecvent în discursul scris al vorbitorilor nonnativi de RLS, întrucât acestea nu au putut fi generate automat.

Prezentăm, mai întâi, cuvintele și expresiile care fac referire, prin forma lor, la discursul oral al vorbitorilor nativi, structuri învățate, în opinia noastră, "după ureche", auzite, cel mai probabil, pe stradă, în căminele studențești, sau în general, în comunicarea cu un vorbitor nativ, sau transferate, din discursul oral în cel scris. Observăm că, de cele mai multe ori, acestea apar corect scrise sau cu mici erori față de formele orginale, fapt ce dovedește memorarea lor ca atare datorită expunerii la un tip de input cu caracter oral. Mai mult, ele apar bine integrate în context, păstrând nota de oralitate pe care o posedă aceste structuri în discursul nativilor. Enumerăm, în continuare, structurile ce s-au evidențiat în corpusul de producții scrise ale vorbitorilor nonnativi. Nu sunt prezente într-un număr mare, însă utilizarea lor la acest nivel ne-a determinat să le inventariem și să le prezentăm aici.

În scrierea primului text, cu tipar interacțional, unii studenți integrează în producțiile lor formule fixe de salut, precum: bună dragule, vă pupic, Pupici!, Eu te invit la mine, Te aștept aici, Dragusta (probabil drăguța). În locul formulelor de încheiere pe care le cunosc (sau se presupune că le cunosc) la acest nivel (Cu drag, de exemplu), apar: Toate cele bune, prietenul tău <<Numel>> sau cel mai bune, <<Numel>>, acest din urmă exemplu putând reprezenta și o traducere literală după expresia englezească All the best. Pe lângă acestea, apar și enunțuri întregi reproduse din discursul nativilor: Şi tu cum te simt? Sper câ te simt bine și tot e în Regula. Astept o scrisoare din partele tau., Apoi dorm-să sperăm ușor și bine, El este drag meu și întotdeauna o se fi. În aceeași categorie, se remarcă expresiile colocviale memorate: duup aia, provenită, probabil, de la dup-aia (după aceea), Acuma, de miază, SEMANAcu, mam deprins cu căldura, dă să fac curățenie, probabil din expresia tre(tră) să fac ceva, ce mai simpatică fată în ..., o prietena de-a mea, știu

din copilărie, nu are teme de nimic, M-am auzit cu colegele, are burtă, ioi!!, jucăm alceva, seara fac de tot, Trebe sa cumpar o banca la calcolator dar nu am unde să put. Tot aici se remarcă folosirea adverbului atunci, cu valoare consecutivă, preluat din discursul oral: Atunci, ne întâlnim anul viitoare!, atunci este departe.

## 7.7.1. Structuri specifice nivelului A1

După cum am văzut și din analiza FA și R1, prepoziția *cu* este una dintre cele mai prolifice prepoziții la nivelul A1. Am investigat contextele de utilizare ale acesteia în corpusul de producții scrise ale vorbitorilor nonnativi și am evidențiat structurile specifice nivelului A1, din punctul nostru de vedere, pe care le prezentăm în continuare:

## Cu + persoană

Am identificat în corpusul nostru 192 de apariții ale acestui grup de cuvinte. Cel mai frecvent, cu apare în prezența substantivului comun prieten: cu prietenul (dar și cu prietena, cu prietenii, cu prietenele sau chiar însoțit de adjectivul pronominal posesiv meu, mea, mei, mele. De exemplu, eu nu pot ma intalnesc cu prietenii mei aici, În weekend el joacă fotbal cu prietenul lui. Apar câteva situații și cu niște prieteni, de pildă: Apoi eu merg la bar cu niște prieteni., am întălnit cu niște prieteni aici. Am putut observa că, de cele mai multe ori, acest tip de structuri sunt folosite corect, ceea ce arată că este vorba despre o structură specifică nivelului A1. Pe lângă aceste situații de ocurență, prepoziția cu apare și în combinație cu alte substantive cu referent persoană, de exemplu: coleg, verișor, mama, tata etc., nume, sau în prezența unui pronume personal, cu ei, cu ele, cu tine (totul ok cu tine?).

## *Cu* + *obiect* (*mijloc de transport*)

O altă structură ce caracterizează nivelul A1 este reprezentată de prepoziția cu în prezența unui substantiv comun care exprimă un mijloc de transport: cu autobuzul/maşina/trenul/avionul/bicicleta. Am observat că, de cele mai multe ori, această structură apare scrisă și folosită corect în context, existând foarte rar omiteri ale articolului definit. Și în acest caz, concluzionăm că este vorba despre un element specific nivelului A1, însușit corespunzător.

Am identificat în corpus și grupurile de substantive conectate prin prepoziția cu, care, prin modul de scriere și de utilizare, ne determină să afirmăm că au fost memorate drept grupuri fixe. Ca dovadă, de cele mai multe ori, acestea sunt corect

scrise și utilizate. Prezentăm, în continuare, un inventar al acestor grupuri: plăcinte cu ciocolată, cafea cu lapte, sandvici cu carne de pui, supă cu legume, sarmale cu smântână, orez cu pui, cereale cu lapte, spaghete cu pui, carne de pui cu legume, pâine cu unt.

O situație similară cu cea prezentată anterior caracterizează și substantivele conectate prin prepoziția de, foarte des întâlnite în corpus. Prezentăm un inventar al acestor structuri: sală de sport (apar și situații de omitere a prepoziției), sală de fitness, club de sport, tenis de masă (situații și cu prepoziția PE: tenis PE masă), facultatea de litere/de chimie/de studii europene etc., casa di cultură, supă de pui, temă de casă, supă de cartofi, suc de mere, carne de porc, carne de pui, salată de verza, friptură de pui, ciorbă de pui (scrisă corbă sau cerbă, cioră), salată de legume, carne de vită, sucul de portocal (suc de portocate), piept de pui, supă de legume, șuncă de poi, salata de vinete, film de acțiune (și filme de acțione), salată de fructe, ziua de naștere, ochelari de vedere, fructe de mare, ochelari de soare, costum de baie, umbrelă de soare, muzeu de animale/istorie/artă, cremă de soare, sac de dormit, camera de zi, produse de baie, fotografii de familie, meci de fotbal, gustare de pui, prajitura de ciocolată, suc de fructe, înghețată de ciocolată, tort de ciocolată, camera de cămin.

Deși nu sunt într-un număr mare, am inventariat și alte grupuri de cuvinte care apar în producțiile scrise: soare cu dinți, apă plată, apă minerală, apă carbogazoasă, fast-food, jocuri video, vază cu flori, telefon fix, cartofi prăjiti.

O categorie aparte o reprezintă structurile de tip *verb* + *prepoziție*: *merg LA/plec LA/mă duc LA*. Întrucât, din experiența noastră didactică, cunoaștem că selecția prepoziției potrivite cu un anumit verb poate reprezenta o problemă, mai ales, pentru vorbitorii de la nivelurile elementare, am încercat să investigăm asemenea structuri. Astfel, am observat că, cel mai frecvent, acestea sunt folosite corect, dar credem că și ele au fost memorate cu prepoziție cu tot, după cum au apărut în input. Afirmăm acest lucru bazându-ne pe apariția în exemple tocmai a acestor structuri și în alte contexte: *merg LA centru/merg LA parc*. De exemplu, *Am merg mulți loci, merg la vivo moal, la munte, Merg la stație, Merg la clasă*.

Astfel, structura *merg ÎN* apare de 26 de ori în corpusul scris, dintre care cele mai multe apariții corecte (25) sunt folosite împreună cu: *oraș, centru, parc. Merg LA* este mult mai frecventă, apărând de 349 de ori în corpus, din care, doar de trei ori, apare *merg LA parc*. Prin urmare, în general, este utilizată corect. Structura *merg CU* are 24 de ocurențe, corect utilizate (vezi **cu** + **obiect**). Având în vedere confuzia *în/la*, am căutat să vedem în ce măsură este corect utilizată și structura *în centru*. Am

observat că, în 21 de cazuri, apare *ÎN centru* și, doar în 9 cazuri, avem o formă greșită, *LA centru*. Așadar, și această structură este una bine stăpânită la nivelul A1.

Pe lângă verbul *merg*, am investigat și verbele *a vorbi* și *a se uita*, întrucât și acestea presupun selecția unei prepoziții în comunicare și știm că nonnativii pot întâmpina probleme în utilizarea lor. Astfel, *vorbesc CU/DESPRE/LA/ÎN* apare în 30 de situații, toate folosite corect, rar cu omiterea articolului la substantiv (cu două excepții: *vorbesc pentru TaTa* și *vorbesc de tată meu*). De exemplu: *vorbesc la telefon cu familia, vorbesc cu prieten/vorbesc cu prieteni mei, vorbesc cu puține cuvinte, Vorbesc cu familia mea la telefon, vorbesc în limba română, vorbesc despre sora mea.* Și verbul reflexiv *a se uita* apare în structura: *a se uita LA televizor* sau *a se uita LA un film*, în 68 situații, majoritatea corecte: *stau acasa ma uit la televezor*, *Sau stau acasă și mă uit la un film.*, *ne uităm la film la sufragariă*. Totuși, apar erori, precum: omiterea pronumelui (*uit la televizor*, *când eu termin uit la film*) sau folosirea formei participiale a verbului (*mergem la mall sau cinma uitat un film*), însă aceste situatii sunt rare.

Din analiza corpusului scris, am observat că sintagmele posesive apar de cele mai multe ori corect scrise, în special, când este vorba despre persoane. Astfel, sintagmele posesive corecte, cu acordul de gen și număr făcut, chiar dacă, uneori, lipsește articolul definit, sunt următoarele: *fratele meu, mama mea, familia mea, camera mea, colegii mei, prietenii mei*. Din punctul nostru de vedere, aceste sintagme sunt și ele memorate. Acestea reprezintă structuri care apar foarte frecvent atât în input, cât și în output, și de aceea studenții le cunosc foarte bine. Ca dovadă, atunci când apare un cuvânt mai puțin frecvent, acordul nu e făcut corect: *apartamentul mea, cameră meu* etc.

O altă structură frecventă în corpusul de producții scrise o reprezintă (în) fiecare zi (apar și cazuri cu fiecare seară, fiecare dimineață, însă acestea sunt mult prea rare, în comparație cu expresia compusă cu substantivul zi). Considerăm că și acestea constituie niște expresii memorate, mai ales că, rar, se omite prepoziția în (în 78 de cazuri dintre cele 318). După cum spuneam, în fiecare seară și în fiecare dimineață apar mult mai rar, doar 11 ocurențe pentru în fiecare seară, corecte, cu excepția unei situații: fiecare seara prieten mea și eu, respectiv, în fiecare dimineață, 16 rezultate, dintre care, în două situații, e omisă prepoziția: Eu mă trezesc ora șapte jumate fiecare dimineața, A fiecare dimineață mănânc Poate salata de vinete.

Am identificat în corpus și formulele fixe de salut, despre care, de asemenea, credem că sunt memorate. Prezentăm, în continuare, cele mai folosite forme, așa cum apar scrise în cele două corpusuri: *Te pup/Vă pup* (21 de apariții, corect utilizate), *Ne vedem* sau *Nevedem peste șase zile, nividem joi* sau chiar *Nevedem pe* 

curent (20 de ocurențe). Scrierea împreună a pronumelui și a formei verbale ne face să concluzionăm că așa au fost memorate aceste forme. Alte formule sunt: Pe curand (12 ocurențe), ne vedem curând! (2 situații), ne întâlnim (10 situații), la revedere (dragă/Drăga prietena/cu drag), 10 ocurențe, Bună ziua (9 ocurențe), Cu mult drag! (7 situații), Servus (2 situații), Bună seara (2 ocurențe), Bună! (o situație), Pe mâine (o situație), o zi bună (o situație), Te salut (o situație)/Te salutat si te iubesc (o situație)! Te iubesc (2 ocurențe), La revedere, pa! (o situație), Ciao! (o situație)/Ceao (2 ocurențe), Cu dor (o situație). Diversitatea formulelor folosite, precum și numărul acestora considerăm că se datorează și cerințelor de la primul subiect de scriere. Influența acestora se va putea observa și mai jos în cazul altor formule fixe, memorate, și folosite ca atare în producții, situație absolut firească la nivelul A1, după cum am evidențiat în Tabelul 11, în grila de evaluare pentru producerea scrisă, din Companion Volume: "Arată un control extrem de limitat asupra unor puține structuri gramaticale și propoziții model, dintr-un repertoriu memorat."

În ceea ce privește alte formule fixe, utilizate de obicei în textele cu tipar interacțional, am constatat că, cel mai des, sunt folosite perechile singular/plural, memorate și ele drept enunțuri complete: Ce mai faci?/Ce mai faceți? (80 de utilizări) și Ce faci?/Ce faceți?, cu 60 de utilizări. Următoarea ca frecvență este propoziția Eu sunt (foarte) bine, cu 35 de ocurențe, dar și Sper că ești/Sper că sunteți bine., cu 20 de ocurențe. Mai puțin frecvente sunt: Cred că (13 situații), cum ești (6 utilizări), cum sunteți (3 ocurențe), mulțumesc (4 situații), Mă bucur (3 situații)/sunt fericită că (o situație). Îmi pare bine (o situație), Îmi pare rău (două ocurențe), vă rog (o situație), Din fericire, (1) sau și mie!, folosită o singură dată, greșit, în loc de construcția și eu.

Dacă, mai sus, am evidențiat propoziții memorate, în continuare, am extras din corpusul scris fragmente (chunks) care sunt clar memorate, în comparație cu restul textului, întrucât acestea sunt fără greșeală sau cu foarte puține greșeli. Cel mai probabil, acestea sunt preluate din exemplele din manual sau din cele oferite de profesor. Am identificat fragmente despre vreme: Timpul e minunat., Vremea este perfectă., Cum e vremea momentan?, Vremea este excelent, despre oraș: Orașul e incridibil., Clujul este un oraș frumos!, Orașul este frumos (cu 5 ocurențe), despre viața în România, în general: Îmi place mult aici., Sunt foarte mulțumită aici., Mă pregatesc pentru un examen de limba romănă., Eu am mulți prieteni foarte buni., Seara, nu fac nimic dar câteodată mă plimb în parc., Ma adaPtez rePede cu viața din Romania., Mă bucur că voi veniți la Cluj curând!, sau formulări folosite, în general, la începutul sau la finalul textelor cu tipar interacțional: Ne vedem când mă întorc., Vino la mine., Abia aștept să te re văd., Ce faceți in weekend?, Sper să ne

întâlnim în vacanta., Sper să ne întâlnim după weekend., Sper sa te voi întălni vinerea viitoara., Tu ce ai facut azi?, Nu te-am mai vazut de foarte mult timp., Ce faci în weekend?, Ne întâlnim în vără?, Voi sunteți ocupați?, sunați-mă când aveți timp., Cand vin acasa abia astep se te vad, al tau., Ne vedem în septembrie, Hai la mine rapede, Hai la o bere cu mine în weekend!, sau alte expresii despre oameni, locuri, obiceiuri: El cel mai mare din frați lui., El ajută mă in totdeauna cind eu am o problema., El pregătește lecțiile pentru elevii lui., Data vitoare vi cu mine, pentru că e distracție., Este foarte trist că nu există un astfel de restaurant., plin de oameni politicosi, Are o vedere despre mare., Dupa cum știi, Noi bem o sticlă de vin franțuzesc, Sună cesaul desteptator, Mă bucur că avem o relație atât de bună., El mă ajută când mi-e greu., Suntem foarte apropriat, ne ajutăm întotdeauna. Un alt fragment care este clar memorat după modelul oferit de profesor îl reprezintă structura nici..., nici, este potrivit(ă)/așa și așa, în funcție de modelul oferit de profesor, folosită la descrierea de persoane: Mama mea nici înaltă nici scundă, e potrivit. sau: Nu este nici slab, nici grass, este potrivit, Ea nu este nici foarte înaltă, nici foarte scundă. El este potrivit, nu este nici inălt, nici scurt, Este nici grasă nic slabă e aşa și aşa., Ea nu e nici grasă și nici slabă, e aşa și aşa. În fine, structura nici..., nici... pare a fi una specifică nivelului A1, fiind găsită în 30 de situații în corpusul de producții scrise.

În ceea ce privește sintagmele fixe, cu prepoziție, care sunt, în general, predate la nivelul A1 (a avea nevoie/a avea chef DE), observăm că ele sunt structuri utilizate frecvent și, de obicei, corect (apar uneori omiteri ale articolului), în special în combinație cu un substantiv: Azi, am nevoie de orez și ouă., Eu am nevoie de o ceai!, Apoi eu am nevoie de puțină odihnă., nu am nevoie de nimic., La ora nouăprezece am chef de o intâlnire cu prietenii mei, Am chef de o dulap cu mancare pe peretele de dreapta. Am constatat că, desi nu constituie o structură specifică nivelului A1, a avea nevoie SĂ pare a fi necesară, întrucât, de multe ori, se folosește de în loc de să: am nevoie de studiat limba rumena, Am nevoie de dormire, am nevoie de odihnat. Considerăm că ar fi util de prezentat contextul de utilizare și pentru această structură pentru a evita fixarea sa într-o formă greșită. Situația e similară și în cazul construcției a avea chef SĂ, remarcând că, în multe situații, această structură apare incompletă sau, în locul ei, este folosită construcția cu DE: Am chef de să găsesc o masă nouă, prentru că chef să mănânce o salată de fructe, are chef să cantă!. Construcția a avea grijă de nu pare a fi specifică nivelului A1, având în vedere că am identificat-o doar în două producții, în ambele situații prepoziția fiind omisă: are grija toata acasa, AI grija tine. Pe de altă parte, observăm că demonstrativul de apropiere (acest, această, aceasta, asta) este prezent în lista noastră, deci și în vocabularul vorbitorilor nonnativi de limba română, cu toate că, din punct de vedere gramatical, pronumele și adjectivul demonstrativ se introduc abia la nivelul A2. Având în vedere FA a acestor elemente, asta = 27, acest = 25, aceasta/această = 19, considerăm că ar fi util și productiv să se ia în considerare includerea lor, în mod organizat, încă de la A1.

Am remarcat o utilizare frecventă a sintagmei *LA Facultatea de Litere*, aceasta apărând în 80 de producții, situație firească, având în vedere, pe de-o parte, statutul de student al subiecților, iar, pe de altă parte, tema textelor (*O zi obișnuită de școală*, de pildă).

Foarte frecvent și corect, în general, sunt folosite și construcțiile: de la... (până) la..., din... până în... sau de... până... Am identificat structura de la... (până) la... în 62 de producții cu referire la oră, acesta părând singurul context de utilizare a acesteia la nivelul A1: am curs de la ora 13 până la ora 17, construcția din... până în... fiind destul de rar folosită (sub 10 utilizări) cu referire la luni și ani (Sunt din aici aici din de septembrie până la februarie., Sunt în Cluj din octombrie până în August), în vreme ce de... până... e mai frecvent folosită (16 ocurențe) pentru a prezenta programul săptămânal: Am fost aici de luni până vineri., Eu stau aici trei zile: de joi până sâmbătă. Prepoziția compusă până la apare și singură în context cu ora: Studiez până la 24.00 și apoi dorm., stau în fatul [pat] până la opt și sfert., azi mergem la facultatea până la oră 5. Toate aceste situații ne arată buna însușire a acestor prepoziții, pe de-o parte, dar și apartenența la nivelul A1 a contextelor menționate.

O situație specială ocupă prepoziția după, locuțiunea conjuncțională după ce, locuțiunea adverbială după aceea și adverbul apoi, care sunt predate studenților străini, în general, împreună ca mijloace de asigurare a coerenței și coeziunii textului. Din analiza noastră, am constatat că prepoziția după este folosită corect, intrapropozițional (După duș gătesc ceva sănătos), însă, de multe ori, este utilizată și drept conector interpropozițional, pentru a înlocui locuțiunea adverbială după aceea (Eu merg la restaurat si mânănc mâncăre tradețional. După eu alerg acasă și studiez, Mănânc cartofi cu carne di pui și beau o apă plată. După merg la casă și fac un duș, În fiecare zi, cand eu am școala, duc copilul meu la gradinița, dupa ce, merg la facultatea de litere.). Și locuțiunea conjuncțională după ce este utilizată, în general, în locul celei adverbiale: De luni pană vineri, eu mai trezesc de la 07:30 apoi eu manânc micul dejeune și cereal după ce eu mai duc la facultatea, Acum sunt la școală După ce eu merg la shopping cu prieteni mea, apoi mergem în parc., Noi avam culsuri ore de la 9 pană la 13. După ce noi mergem la camera. Totuși, atunci când este utilizată, locutiunea adverbială după aceea este folosită corect. Constatăm,

așadar, din aceste observații, că locuțiunea conjuncțională *după ce* nu este însușită la nivelul A1, iar locuțiunea adverbială *după aceea* pare a pune unele probleme în utilizare. În ceea ce privește adverbul *apoi*, situația pare a fi una bună, acesta fiind folosit, de cele mai multe ori, adecvat: *Apoi mă duc la baie, mă spăl și, după aceea, ies camera mea.*, dar, și în acest caz, apar confuzii cu prepoziția *după*, în ceea ce privește contextul de utilizare: *apoi anul pregatitor se studieze arhitectura*. Luând în considerare situația descrisă, din punctul nostru de vedere, nu toate structurile cu *după* caracterizează nivelul A1. De aceea suntem de părere că locuțiunea conjuncțională *după ce* caracterizează mai degrabă nivelul A2, și nu nivelul A1.

Pe lângă cuvintele de mai sus, am identificat în corpus preferința pentru o serie de conectori. Astfel, se utilizează extrem de des conjuncția pentru că (în 325 de situații): Serios, îmi place, pentru că ea este mama mea., După aceea, mă uit la televizor pentru că îmi place să odihnesc acasă singură., Ea știe foarte mulți informații pentru că ea e studiază mulți. De asemenea, se folosesc frecvent, dar și corect, locuțiunile adverbiale de obicei (115 ocurențe) și în general (10 situații): De obicei merg acasă cu prietene mele, După curs, în general mânănc cu prieteni mei în restaurant lângă căminul mea. Utilizări mai puține au locuțiunea adjectivală în plus (6 situații) și cea adverbială, în final: El prefară Să citire, a final mă culc pe la ora 11, cu doar 3 ocurențe. În cazul ultimei locuțiuni, am constatat că aceasta nu este folosită corect, utilizându-se de fiecare dată o formă cu prepoziția la: la final.

Locuțiunea adverbială de exemplu are 13 ocurențe în corpus, apărând în multe texte corect, însă, de multe ori, se omite prepoziția: În camera mea nu îmi place de multe obiectele exemplu calculator, Ea poată să gătește bune mâncare, Exemplu: pizza, supa de legume., Exista multe si deferet mancațiexemplu sarmalele. Am identificat și forme cu la sau par (la exemplu sau par exemplu), probabil traduceri literale din engleză (for example) si franceză (par exemple).

Și locuțiunea adverbială pe jos apare în 18 situații, dintre care 17 sunt în structură cu verbul a merge, fiind folosite corect. Pe jos, cu sensul de "podea", apare doar o dată în exemplul: un covor roși pe jos. Locuțiunea adverbială de asemenea este utilizată în 18 situații: Noi suntem de asemenea prieteni din cameră la caminu., El vorbeste de asemenea limba franceza și îngleza., În fiecare zi mergem în oraș și schiam de cinci ori pe zi. De asemenea vizităm musee și catedrale.

Locuțiunea adverbială un pic are 16 ocurențe: Apoi mă culc un pic, Eu sunt obosit pentru ca eu am un pic bolnav. discut un pic cu <nume1>, este un pic dificile pentru mine, Ei este un pic mai înalt, este un pic de soare.

Contrar așteptărilor noastre, structuri ca: *încă o dată, zi de zi, o dată pe lună, de două ori pe lună, departe de apar* foarte rar, până la 3 ocurențe: *O dată pe lună* 

ne întâlnim și bem cafele și măncăm biscuiți., El merge la cinema de două ori pe luna. Locuțiunea prepozițională aproape de nu apare corect reprezentată în corpusul scris, însă am constatat necesitatea ei, în enunțuri de tipul: Este foarte mare, comfortabilă și comodă, aproape e Centru. Atunci, nu îmi place ceva despre bars, clubs sunt aproape la camin, O ciuvetă este lângă perete este aproape ușă. Apare și o variantă a acesteia, apropiat de: un mic oraș și apropiat de Verona.

Remarcăm o nesiguranță în utilizarea cratimei la cuvinte precum după-masă/după-amiază: după masă eu manănc în la Cantina, În fiecare după masă mergem la plembare, de obicei după masă ma duc la plaj, am cursuri de la ora 9 până la ora 1 dupa masă, de obicei dupa masă, eu mă uit la stradă., dupa masă eu si colegii mei de camera mergem impreună în centru, Dupa masa eu merg acasa și mananc pranz, de la oră nouă dimineață până la oră unu după amiază, Dupa-amiaza eu fac temă, După-amiază eu merg la curs.. Pe de altă parte, în cazul substantivului mic dejun (în corpus și Micul dejun/mecul dejun) apare uneori folosită cratima. Remarcăm că acesta este folosit, în general, corect, cu verbul a lua sau în structura mănânc la micul dejun. Desigur că apar și formulări ca: mănânc/pregătesc/gătesc micul dejun.

Expresiile stării, așa cum le sunt prezentate studenților străini verbele cu pronume în dativ, *a-i fi foame, a-i fi sete, a-i fi rău, a-i fi dor, a-i fi frică* etc., nu sunt pe atât de des utilizate pe cât am fi crezut. Doar *mi-e dor de...* apare în 38 de situații, dintre care cele mai multe sunt în formularea *mi-e dor de tine!*, structură clar memorată. În opinia noastră, unele dintre aceste verbe nu caracterizează nivelul A1. Totuși, propunem păstrarea în inventarul lexical de nivel A1 a următoarelor: *a-i fi foame, a-i fi sete, a-i fi somn*. În aceeași linie se înscrie și verbul *a iubi*, care este introdus, conform *Descrierii minimale*, la nivelul A2, însă, după textele studenților, acesta este necesar încă de la A1, mai degrabă în formulări fixe, ce vor fi apoi memorate (situație ce se întâmplă deja), nu întreaga paradigmă: *Te iubesc, Vă iubesc, il iubesc tatal meu. Te iubesc mamă. Eu o iubesc pe sora mea. Eu îl iubesc pe tată.* Îl Iubesc foarte mult pe frate meu, Eu îl iubesc pe Tatal meu Foarte mult.

În analiza noastră, am observat că locuțiunea tot timpul apare destul de frecvent în producțiile nonnativilor, în general, corect: El ascultă muzică tot timpul., poartă ochelari tot timpul., tot timpul, vrem să mergem cu tatăl meu., Ea este amabil, generoasă, inteligentă și serios dar nu tot timp. veselă și zambeste tot timp. Pe de altă parte, colocația timp liber, când se află în combinație cu verbul a avea, apare întotdeauna corect, deci e caracteristică nivelului A1: Când am timp liber pot să merg în parc sau la cinema, Ea nu are mult timp liber pentru că studiază mult. Cu prepoziția în, apare și nearticulat: În timp liber eu alerg în parc, În timp libere meu

eu vizitez multe orași, joacă Tenis in timp liber, în timpul liber, eu se uita la televizor, merg în parc., În timpul liber, de obicei, citesc o carte, ascultat muzică și fac plimbare în parc.

Am observat că sunt confundate verbele *a cânta* și *a juca*, cel mai probabil din cauza traducerii din limba engleză unde se face diferența între *sing* (cântă cu vocea) și *play* (cântă la un instrument, dar și joacă un joc/un sport). Astfel, întâlnim în producțiile nonnativilor exemple precum: *cântă la flute.*, *El cântă 5 la instrumente de muzica și face muzic*, *cântă pe guitar*, *În timpul liber cântă*, *joacă pianul*. De observat că, în acest ultim exemplu, sunt folosite ambele sensuri ale verbului *a cânta*. Totuși, nu pare a exista aceeași problemă în cazul verbului *a juca*, acesta fiind folosit corect: *Eu joaca fotball cu prietenii în parc*.

Am încercat să evidențiem și strategiile utilizate de către studenți în situațiile în care au nevoie de un anumit element lexical, dar nu dispun de acesta. Adesea sunt folosite strategii de aproximare a sensului, prin descriere sau parafrazare: are par scurt și albi și negru împreună. (= cărunt), Nu este doctară, dar ajuta oameni care nu sunt bine (nurse) (= asistentă medicală), pantofi de munte (= bocanci), camera de dormi (= dormitor), Învață limba Bretonne, limba mama ei (= limba maternă)., Mergi tot sâtămana Pe strada (= a face mișcare, a alerga) și mănance Fructe Pentru A Fi sanatoasa. Sunt situații când se folosește un alt cuvânt similar ca sens, însă inadecvat în context: Iar nu ima place sertarul mea pentru ca este in virsta (= vechi)., Peste anul Pregatitori eu merg la spital in Centru. (= după)., Mama mea este nici tanâră, nici vechă (= bătrână); ea are 50 ani, sau chiar verbul în locul substantivului: costa de inchere, Dar costă e mic. Alte strategii utilizate mai rar sunt: traducerea literală – "Lacul de Podea" (Bodasee, în germană, Boden = podea, iar See = lac/mare) – sau transferul din limba engleză, Toțe obiectele sunt în order.

O altă dovadă a faptului că, la acest nivel, se memorează de cele mai multe ori structurile o constituie și exemplele de mai jos: *Şi, și mie și lui <Nume1> Atât mie, cât și*, studentul a folosit ambele structuri corelative, așa cum se obișnuiește să fie predate la clasă. Aceeași situație o regăsim și în exemplul *de la 9:00 la (până) 13:00*, aici prepoziția fiind dislocată, sau în enunțul *Nevedem pe curent*.

Totuși, am remarcat că unele teme par a fi mai potrivite decât altele. De exemplu, în cazul descrierii de persoane, am constatat că se folosesc mult mai puține formulări memorate, iar studenții par a fi mult mai motivați să producă singuri discurs.

## 7.8. Criteriul întregirii sferelor lexicale

După cum am arătat în partea întâi a lucrării noastre, în stabilirea listelor de cuvinte, pe lângă criteriile statistice, se folosesc si unele mai mult sau mai putin subjective, cum este cazul criteriului întregirii sferelor lexicale. Asa cum am specificat în cazul substantivelor proprii, am decis ca, și în această situație, să evidențiem cuvintele care fac parte din aceeași sferă lexicală și care se învață, în general, împreună. Pe lângă cele regăsite în listă (si pe care le-am păstrat acolo), unde am considerat necesar, am adăugat unul sau mai mulți membri pentru acea sferă lexicală. Am avut în vedere 14 teme: culori, părți ale corpului, timp, momente ale zilei, anotimpuri, zilele săptămânii, lunile anului, numerele, grade de rudenie, animale, politete, sporturi, mijloace de transport si obiecte din clasă. Nu am adăugat toți membrii unei sfere lexicale, încercând, pe cât a fost posibil, să nu includem un număr prea mare de cuvinte, ci, mai degrabă, cuvintele care, în opinia noastră, ar putea fi necesare în această etapă din diferite motive. De exemplu, în cazul culorilor, am adăugat doar culoarea moy, însă, în ceea ce priveste lunile anului, ne-am asigurat că am inventariat toti membrii grupului. La fel am procedat și în cazul momentelor zilei. În alegerea acestor elemente, ne-a ajutat experienta didactică (care sunt cuvintele cerute, în general, de către studenți, cuvintele internaționale sau acelea care sunt necesare si în discutarea unor probleme gramaticale). Pentru a identifica mai ușor cuvintele adăugate, am optat pentru marcarea lor în lista de mai jos cu culoarea gri.

Tabelul 19 – Lista sferelor lexicale

Sfera:	Culori		Părți ale corpului		Gradele de rudenie		Animale	
Nr. crt.	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA
- 127 5231	negru	192	ochi	144	mamă	313	pui	119
	roșu	82	păr	181	tată	238	câine	34
	alb	70	dinți	20	copii	203	pește	19
	albastru	61	ureche		soră	95	vacă (vită)	18
	verde	46	gură		frate	93	porc	14
	culoare	25	nas		bunic, bunică, bunici	21	animal	13
	maro	22	față		fiu	16	pisică	10
	galben	14	gât		părinți	15	oaie	10
	gri	11	mână		sot, sotie	13	elefant	
	T	11	picior		mătusă	4	girafă	
	roz portocaliu	7	deget		unchi	4	cal	
		/			fiică	4		
	mov		cap				urs	
					nepot verisor			
					,			
Sfera:	Momente		Anatimay		verișoară Zilele		Millagas da	
Siera:	ale zilei		Anotimpu ri		săptămânii		Mijloace de transport	
Nr. crt.	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA
INI. CIT.	seară	169	vară	75	luni	66	mașină	106
	dimineață	130	iarnă	41	vineri	38	avion	84
	noapte	16	primăvară	22	duminică	27	autobuz	84
	după-masă	12	toamnă	4	sâmbătă	22	tren	46
	după-amiază	12				22		
	amiază		anotimp	4	joi	13	bicicletă	16
					miercuri	8	taxi	12
	1				marți	4	motocicletă	6
							tramvai	
Sfera:	Timp		Lunile anului		Politețe		Sporturi	
Nr. crt.	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA
	oră	613	septembri e	8	mulţumesc	101	fotbal	151
	zi	417	iulie	7	mersi	9	sport	68
	an	397	octombrie	7	cu plăcere	29	înot	58
	timp	188	februarie	5	te/vă rog	58	tenis	33
	săptămână	108	august	5	mă scuzați/scuze	28	volei	21
	lună	50	decembrie	5	poftim	13	schi	19
	mâine	49	ianuarie	4	poftiți	44	baschet	10
	jumătate de oră	43	iunie	4	dumneavoastră		şah	10
	azi	36	noiembrie	4	domn		alergare	
	minut	15	martie		doamnă		gimnastică	
	astăzi	15	aprilie					
	ieri	14	mai					
-	sfert de oră	9						

	fix	37						
	secundă							
Sfera:	Obiecte din clasă		Numere		Numere		Numere	
Nr. crt.	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA	Cuvinte	FA
	carte	183	zero	162	cincisprezece	13	milion	
	birou	99	unu, una	255	şaisprezece	15	număr	43
	caiet	65	doi, două	676	şaptesprezece	15		
	sală de clasă	37	trei	334	optsprezece	34		
	pix	19	patru	245	nouăsprezece	17		
	tablă	10	cinci	191	douăzeci	111		
	creion		șase	88	treizeci	74		
			şapte	99	patruzeci	8		
	•		opt	128	cincizeci	26		
			nouă	62	șaizeci	11		
			zece	40	şaptezeci	4		
			unsprezece	86	optzeci	4		
			doisprezece	31	nouăzeci	6		
			treisprezece	32	sută	24		
			paisprezece	68	mie	7		

### 7.9. Subcriteriul necesității

Așa cum am subliniat în capitolul 5, este important să avem în vedere și vocabularul necesar pentru înțelegerea cerințelor în cadrul cursurilor și al examenelor. De aceea am elaborat un mic corpus de cerințe (prezentat în capitolul 6), pe baza căruia am extras elementele de vocabular esențiale pentru înțelegerea sarcinilor la nivelul A1. Considerăm că și aceste elemente, mai ales în primele zile de curs, intră în vocabularul studenților ca formule memorate, ceea ce este specific nivelului A1 (vezi *Tabelul 11*, Grila de evaluare din *Companion Volume*). În tabelul de mai jos, expunem elementele cele mai frecvente din cerințe și care, în opinia noastră, ar trebui să fie incluse și în inventarul lexical al RLS:

Tabelul 20 – Elementele lexicale propuse pe baza criteriului necesității

Nr.	Cuvântul	Colocații	FA în corpusul
crt.			de cerințe
	imagine/imaginea/		127
	imagini/imaginile		
	ce		116
	completați		108
	scrieți		108
	ascultați	ascultați + ceva (= un text, un dialog, niște informații, 5 dialoguri, mesajul, conversația etc.) ascultați + cineva (un client, o persoană, trei persoane) ascultați + citiți/verificați/scrieți/spuneți	104

text/texte	ascultați/citiți/scrieți un text	95
citiți		82
potrivit (adj.)	imaginea potrivită, răspunsul potrivit, imaginea potrivită	75
perechi	Lucrați în perechi!	74
corect (adj.)	răspunsul corect/varianta corectă/literele corecte	65
fiecare	folosiți fiecare text/fiecare imagine	64
alegeți	alegeți imaginea/varianta/răspunsul	62
libere	spații libere	62
decideți	doar în combinație cu A/F	51
adevărat/adevărate		50
fals/false		50
potriviți		48
care		47
spuneți		42
dialog/dialoguri		36
ordonați	puneți în ordine/puneți în ordinea corectă	18
încercuiți		2
informații		1

După cum se poate observa, am optat să păstrăm, în lista de mai sus, formele cuvintelor așa cum sunt ele folosite, în general, în formularea sarcinilor, având în vedere că unele dintre acestea vor fi necesare înainte de a se introduce categoria gramaticală.

Confruntând această listă cu lista generală, observăm că lipsesc din cea din urmă cuvintele: a completa, corect, a decide, adevărat, fals, a potrivi, dialog, a ordona, a încercui, substantiv, verb, adjectiv, articol, prepoziție, pronume.

Un alt aspect de analizat în cazul listelor de frevență este reprezentat de *problema melcului de mare*. În cazul nostru, am optat pentru verificarea acestei situații doar în ultima etapă, presupunând că multe dintre cuvintele aflate sub incidența acestei probleme vor fi excluse în etapele intermediare, ceea ce s-a și întâmplat, în cea mai mare parte.

Tabelul 21 – Excluderea unor cuvinte pe baza principiului melcului de mare

Cuvânt	FA	FR	R1	R2	R3	R4	whelk
shaorma	5	0	1	1	1	1	1
ajutor	6	1	3	1	1	1	2
bilet	6	1	2	2	1	1	2
câmp	6	1	1	1	1	1	2
fotograf	6	1	1	1	1	1	2
reducere	6	1	3	1	1	1	2

Rămâne doar situația celor 6 cuvinte din lista de mai sus. Ne atrag atenția cuvintele care au R1 = 1 și R4 = 1, ceea ce presupune că aceste elemente lexicale apar doar într-un singur text, de 4 ori, cum este cazul cuvintelor: *shaorma*, *câmp* și *fotograf*. Din acest motiv, am decis excluderea lor din inventarul lexical al RLS. Totuși, cuvântul *fotograf* este un cuvânt util, la nivelul A1, pe de-o parte datorită apartenenței sale la clasa profesiilor, iar, pe de altă parte, prin faptul că este un cuvânt internațional și deci ușor inteligibil. Aceeași situație credem că este și în cazul celorlalte cuvinte, care, deși nu au FA mare, sunt relevante pentru acest nivel.

Tot în această etapă, am constatat că, din lista noastră, sunt omise două adjective, cu frecvență redusă, cel mai probabil, însă necesare atunci când se descriu persoane, temă specifică nivelului A1: **roșcat** și **cărunt**, pe care le vom adăuga listei. De asemenea, având în vedere că, în lista de frecvențe, apare substantivul *ofertă*, am introdus și verbul **a oferi**.

După o confruntare a listei cu *Descrierea minimală* pentru a verifica includerea tuturor elementelor necesare din perspectiva gramaticii, am constatat necesitatea întregirii inventarului cu o serie de pronume (*ție, nouă, vouă*), dar și cu verbul auxiliar de perfect compus, *a avea*: **ție, nouă, vouă, am, ai, a, ați, au**.

Întrucât fiecare cuvânt a apărut sub diferite forme în lista de frecvențe, în final, am optat pentru reținerea în inventarul lexical a formei de bază a cuvântului, iar forma în locul căreia apare aceasta, este prima inclusă în coloana de *alte forme* (vezi tabelul de mai jos). Totuși, în cazul cuvintelor *rog*, *pup*, *revedere*, *general*, *obicei* și *exemplu*, întrucât am constatat că sunt folosite în structuri fixe, memorate, am decis includerea lor în inventar, în formele: *te/vă rog*, *te/vă pup*, *la revedere*, *în general*, *de obicei* și *de exemplu*. În cazul altor cuvinte, am optat să nu intervenim în acest fel în listă, deoarece cuvântul a fost folosit și singur în diferite contexte.

În scopul observării diferențelor rezultate în urma etapelor prin care a trecut lista de frecvențe, prezentăm, în tabelul de mai jos, primele 11 elemente inventariate.

#	Cuvânt	Alte forme	FA	FR	R1	R2	R3	R4
1	a fi	e, este, eşte, esti, eşti, sunt, suntem, sunteți, est, fi	5319	445	1888	3804	727	482
2	și	si	5244	438	1039	819	622	447
3	la		3161	264	903	676	485	332
4	în	in, en, ăn	2995	250	1205	778	440	248
5	eu	ieu	2595	217	768	534	390	279

Tabelul 22 – Primele 11 cuvinte din lista finală de frecvențe

6	de	di	2128	178	869	528	301	175
7	a avea	am, ai, are, avem, aveți, au, avea	1971	165	1011	236	252	66
8	cu		1836	154	801	469	226	111
9	meu	mea, mia, mei, melem mieu, miau	1530	128	933	335	142	67
10	0		1497	125	623	324	178	113
11	un		1429	119	546	307	200	130

La o primă vedere, observăm că, acum, primul cuvânt din listă nu mai este  $\hat{si}$ , ci verbul afi, cu FA de 5319. Totuși, considerăm în continuare conjuncția  $\hat{si}$  drept cel mai frecvent cuvânt din corpusul nonativilor de română ca L2, de nivel A1, deoarece, sub eticheta verbului afi, după cum putem observa din coloana *alte forme*, sunt incluse mult mai multe elemente.

O altă modificare notabilă o reprezintă includerea verbului *a avea*, cu FA de 1971, și a adjectivului pronominal posesiv *meu*, FA = 1530, frecvențe obținute prin agregarea tuturor formelor întâlnite în lista inițială.

# 7.10. Legea lui Zipf aplicată listei de frecvențe din corpusul de producții scrise ale vorbitorilor nonnativi de RLS

După cum evidențiam în capitolul 6, numai primele cuvinte dintr-o listă de frecvențe vor avea FA foarte mare, după care procentul ocurențelor va scădea, dar predictibil, conform legii lui Zipf (vezi, în acest sens, capitolul 6). Am verificat și în cazul listei noastre de frecvențe acest principiu și am remarcat existența unui tipar după care se reduc frecvențele cuvintelor în corpusuri, așa cum apare evidențiat în primul grafic.

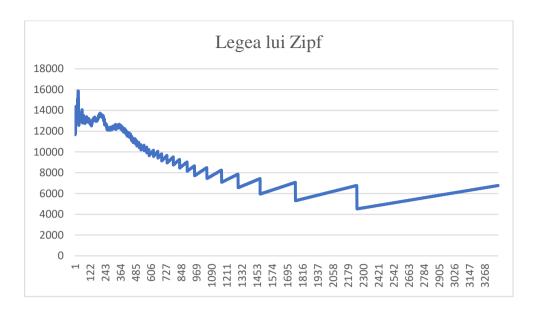


Figura 10 – Legea lui Zipf, Lista de frecvență v2

Dacă privim comparativ cele trei grafice de mai jos, putem observa că, în ceea ce privește FA, în cazul primelor 400 de cuvinte din listă, există o constanță, în vreme ce, trecând mai departe, la primele 500, apare un trend descrescător, care este mult mai evident dacă privim graficul pentru primele 1000 de cuvinte. În acest ultim grafic, se poate observa că, între primele 700 și 800 de cuvinte, frecvența este tot mai mică, iar distanțele pe grafic, tot mai mari. Această constatare reprezintă un factor care ne-a confirmat limita finală pe care am stabilit-o listei noastre, de 760 de cuvinte.



Figura 11 – Legea lui Zipf – primele 400 de cuvinte

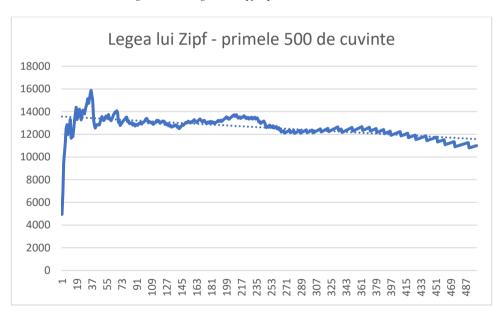


Figura 12 – Legea lui Zipf – primele 500 de cuvinte

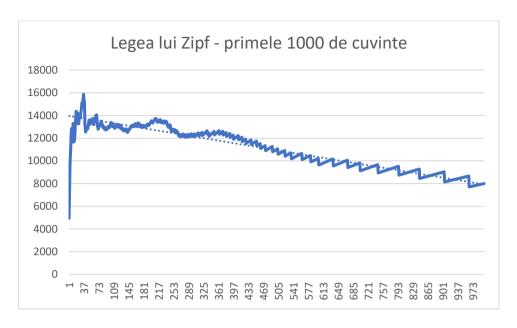


Figura 13 – Legea lui Zipf – primele 1000 de cuvinte

#### 7.11. Concluzii

Am pornit de la o listă de frecvențe de aproape 10.000 de cuvinte, pe care am redus-o în diferite etape, ajungând, în final, la o listă de frecvențe de 757 de cuvinte păstrate din lista inițială. În scopul definitivării inventarului lexical al RLS, pentru nivelul A1, acestei liste i-am adăugat 82 de cuvinte propuse prin prisma criteriilor necesității, a completării sferelor lexicale și a criteriului subiectiv, ajungând la un inventar de 845 de cuvinte. Acestora le-am alăturat lista substantivelor proprii, a abrevierilor și a acronimelor, specifice nivelului A1.

Întrucât ar ocupa mult prea mult spațiu, în corpul lucrării includem doar forma finală a inventarului (însă păstrând ordinea conferită de FA), alături de lista substantivelor proprii, a abrevierilor și a acronimelor (*Tabelul 19*).

Tabelul 23 - Inventarul lexical al limbii române ca L2, nivelul A1

1. a fi	2. și	3. la	4. în
5. eu	6. de	7. a avea	8. cu
9. meu	10. o	11. un	12. pentru
13. merge	14. foarte	15. a-i plăcea	16. nu
17. că	18. da	19. mult	20. cameră
21. zi	22. îmi	23. ce	24. ora
25. pe	26. a mânca	27. mă	28. a face
29. să	30. mare (adj.)	31. din	32. sau
33. bine	34. ea	35. el	36. prieten
37. imagine	38. dar	39. frumos	40. după
41. an	42. a studia	43. facultate	44. fiecare
45. mamă	46. bun	47. oraș	48. lângă
49. aici	50. pat	51. ei	52. noi
53. limbă	54. familie	55. apoi	56. acasă
57. a vrea	58. a bea	59. tată	60. acum
61. persoană	62. curs	63. parc	64. mai
65. a sta	66. română	67. copii	68. masă
69. neagră	70. casă	71. îi	72. restaurant
73. timp	74. mâncare	75. păr	76. carte
77. până	78. tu	79. student	80. fata
81. tot	82. a găti	83. seară	84. dulap
85. a putea	86. munte	87. văd	88. de obicei
89. mic	90. liber	91. fotbal	92. literă
93. a locui	94. se	95. ochi	96. prefer
97. unde	98. birou	99. a purta	100. micul dejun
101. acolo	102. ne	103. vacanță	104. dimineață
105. școală	106. a lua	107. vreme	108. a se trezi
109. fereastră	110. cât	111. când	112. cald
113. pui	114. a lucra	115. raft	116. haină
117. apartament	118. a crede	119. a ști	120. a vorbi
121. a citi	122. bucătărie	123. suc	124. calculator
125. text	126. săptămână	127. a se duce	128. mașină
129. televizor	130. cafea	131. fruct	132. ceva
133. profesor	134. mie	135. mulţumesc	136. lapte
137. a (se) uita	138. a juca	139. soră	140. prânz

141. înalt	142. a dori	143. scaun	144. weekend
145. a se plimba	146. femeie	147. frig	148. soare
149. a costa	150. frate	151. legumă	152. lui
153. muzică	154. salată	155. cum	156. bărbat
157. ok	158. împreună	159. lung	160. după aceea
161. autobuz	162. avion	163. despre	164. film
165. roșu	166. roșie	167. stradă	168. într-un
169. nici	170. a dormi	171. cămin	172. slab
173. joc	174. lei	175. vară	176. baie
177. plimbare	178. carne	179. într-o	180. plajă
181. uneori	182. cartof	183. hotel	184. a alerga
185. alb	186. voi	187. a se numi	188. club
189. puțin	190. scurt	191. sport	192. mine
193. vă	194. luni	195. te	196. mall
197. obiect	198. supă	199. dormitor	200. harnic
201. nevoie	202. fericit	203. inteligent	204. la revedere
205. albastru	206. medicină	207. a călători	208. care
209. salut	210. a veni	211. om	212. pâine
213. înot	214. te/vă rog	215. a (se) spăla	216. apă
217. cinema	218. colegă	219. centru	220. portocală
221. perete	222. a vizita	223. pizza	224. atât
225. cumpărături	226. prăjitură	227. uṣă	228. a asculta
229. a cumpăra	230. a iubi	231. bibliotecă	232. cereale
233. confortabil	234. iar	235. lună	236. mâine
237. pantaloni	238. țară	239. universitate	240. niște
241. cină	242. loc	243. a învăța	244. băiat
245. sănătos	246. tren	247. verde	248. interesant
249. orez	250. canapea	251. dacă	252. numai
253. poftiți	254. elegant	255. jumătate	256. număr
257. scump	258. a se culca	259. cantină	260. ciocolată
261. ieftin	262. duş	263. iarnă	264. în general
265. muzeu	266. nume	267. a se juca	268. bancă
269. covor	270. dor	271. mers	272. sub
273. blond	274. drept	275. generos	276. mare (subst.)
277. singur	278. alt	279. amabil	280. fost
281. nou	282. preferat	283. vineri	284. a se întoarce
285. a spera	286. fix	287. sală	288. azi

289. bere	290. ceai	291. mere	292. poate
293. bluză	294. doctor	295. mi-e	296. a ieși
297. câine	298. ochelari	299. kilogram	300. telefon
301. tenis	302. balcon	303. desert	304. între
305. paste	306. perfect	307. scund	308. tablou
309. temă	310. a pregăti	311. lampă	312. bar
313. examen	314. gata	315. probabil	316. activitate
317. cu plăcere	318. gras	319. tine	320. tricou
321. a ploua	322. a schia	323. a scrie	324. a se întâlni
325. întotdeauna	326. scuze	327. ta	328. a se îmbrăca
329. asta	330. cine	331. clasă	332. duminică
333. jos	334. ou	335. piscină	336. profesie
337. ciorbă	338. deci	339. față	340. floare
341. șaten	342. acest	343. aşa	344. culoare
345. ele	346. fotografie	347. viitor	348. adresă
349. brunet	350. cola	351. cort	352. diferit
353. engleză	354. afară	355. dificil	356. nu-mi
357. petrecere	358. politicos	359. sarmale	360. avantaj
361. elev	362. frigider	363. maro	364. ordonat
365. primăvară	366. sâmbătă	367. a părea	368. aproape
369. banană	370. bani	371. bunic	372. grădină
373. lac	374. normal	375. vin	376. volei
377. a spune	378. calm	379. curat	380. dinți
381. ideal	382. oglindă	383. rece	384. a termina
385. ales	386. altceva	387. căprui	388. fără
389. îți	390. ori	391. perdea	392. peste
393. piață	394. pix	395. schi	396. spital
397. tepup/vă pup	398. un pic	399. vechi	400. a aștepta
401. a se odihni	402. a trebui	403. brânză	404. deexemplu
405. făcut	406. fotoliu	407. hamburger	408. jachetă
409. onest	410. pantof	411. pernă	412. problemă
413. stadion	414. va	415. vă	416. vită
417. a alege	418. a ninge	419. bicicletă	420. delicios
421. des	422. fel	423. istorie	424. lucru
425. noastră	426. organizat	427. tău	428. transport
429. a se relaxa	430. fiu	431. magazin	432. mijloc
433. nimic	434. noapte	435. obosit	436. plată

	438.		
437. senin	supermarket	439. a discuta	440. a pleca
441. a se chema	442. a vedea	443. aragaz	444. astăzi
445. cel	446. curând	447. de asemenea	448. mâncat
449. minut	450. pahar	451. părinți	452. său
453. studii	454. a cânta	455. a suna	456. atunci
457. există	458. farfurie	459. galben	460. ieri
461. iubită	462. minunat	463. nu-i	464. porc
465. sandvici	466. a dansa	467. amuzant	468. an pregătitor
469. animal	470. creativ	471. departe	472. joi
473. m-am	474. meci	475. om	476. ori
477. pătură	478. poftim	479. poză	480. prea
481. prin	482. sănătate	483. soție	484. zahăr
485. a ajunge	486. a folosi	487. a fuma	488. aş
489. avut	490. barbă	491. bucătar	492. chef
493. doar	494. dulce	495. după-masă	496. exact
497. iaurt	498. mereu	499. ondulat	500. pian
501. prima	502. rău	503. roz	504. taxi
505. teatru	506. trecut	507. venit	508. a ajuta
509. alimentar	510. deja	511. devreme	512. euro
513. franceză	514. gri	515. lor	516. minerală
517. niciodată	518. noștri	519. rucsac	520. sticlă
521. terminat	522. tradițional	523. zăpadă	524. a urca
525. această	526. baschet	527. coş	528. dată
529. dentist	530. fast-food	531. idee	532. imediat
533. încă	534. informație	535. le	536. meniu
537. pește	538. pisică	539. prăjit	540. repede
541. şah	542. șofer	543. spaghete	544. tablă
545. alcool	546. calitate	547. castravete	548. clasică
549. copac	550. creţ	551. etaj	552. să fie
553. fiindcă	554. îl	555. important	556. în plus
557. înghețată	558. mersi	559. natură	560. pădure
561. program	562. rar	563. rest	564. sfert
565. sigur	566. sportiv	567. stat	568. vază
569. a descrie	570. aproximativ	571. bej	572. căciulă
573. catedrală	574. chelner	575. costum	576. de acord
577. dreapta	578. european	579. foame	580. greu

581. gros	582. liceu	583. mămăligă	584. mese
585. miercuri	586. nicio	587. obișnuit	588. ofertă
589. politică	590. posibil	591. rapid	592. septembrie
593. serios	594. spălat	595. stânga	596. turistic
597. unt	598. viață	599. ziar	600. zilnic
601. a recomanda	602. a se simți	603. cămașă	604. castel
605. incredibil	606. iulie	607. leneş	608. mână
609. membru	610. naștere	611. nisip	612. noptieră
613. octombrie	614. ocupat	615. pară	616. picnic
617. podea	618. pop	619. portocaliu	620. potrivit
621. rochie	622. salon	623. sos	624. spre
625. vanilie	626. vedere	627. ajuns	628. ajutor
629. bilet	630. câteva	631. citit	632. corp
633. deschis	634. excelent	635. final	636. fotograf
637. hai	638. înainte	639. închis	640. lume
641. minge	642. mulţumim	643. muncă	644. naționalitate
645. nota de plată	646. reducere	647. simplu	648. turist
649. vom	650. acea	651. august	652. băutură
653. bon	654. business	655. cabană	656. caiet
657. cam	658. cea	659. croisant	660. decât
661. decembrie	662. februarie	663. fiert	664. foc
665. fresh	666. fustă	667. ghete	668. grup
669. iarbă	670. inginer	671. masterat	672. matematică
673. nimeni	674. nor	675. pauză	676. preţ
677. relaxat	678. s-a	679. sare	680. serviciu
681. spaniolă	682. sud	683. tort	684. totul
685. văzut	686. veselă	687. video	688. a începe
689. a înțelege	690. a reveni	691. a zice	692. acțiune
693. anotimp	694. arenă	695. biscuit	696. călătorie
697. cap	698. casnică	699. câteodată	700. ceas
701. cer	702. chel	703. chitară	704. client
705. computer	706. curățenie	707. curse	708. deși
709. destinație	710. dezavantaj	711. direct	712. drăguț
713. farmacie	714. geografie	715. gramatică	716. ianuarie
717. înapoi	718. internet	719. italiană	720. iunie
721. laptop	722. lecție	723. luminos	724. marți
725. mătușă	726. modern	727. negativ	728. nervos

729. noiembrie	730. parter	731. pereche	732. plăcut
733. produse	734. radio	735. râu	736. rock
737. roman	738. sef	739. sincer	740. special
741. spus	742. stație	743. stomatolog	744. struguri
745. sunat	746. sus	747. tânăr	748. târziu
749. taximetrist	750. tip	751. toamnă	752. total
753. totdeauna	754. trezit	755. varză	756. vas
757. zis	758. mov	759. cap	760. ureche
761. gură	762. nas	763. față	764. gât
765. mână	766. picior	767. deget	768. fiică
769. nepot	770. verișor	771. verișoară	772. oaie
773. elefant	774. girafă	775. cal	776. urs
777. amiază	778. tramvai	779. secundă	780. martie
781. aprilie	782. mai	783. domn	784. doamnă
785. dumneavoastră	786. alergare	787. gimnastică	788. milion
789. creion	790. zero	791. unu, una	792. doi, două
793. trei	794. patru	795. cinci	796. şase
797. şapte	798. opt	799. nouă	800. zece
801. unsprezece	802. doisprezece	803. treisprezece	804. paisprezece
805. cincisprezece	806. şaisprezece	807. şaptesprezece	808. optsprezece
809. nouăsprezece	810. douăzeci	811. treizeci	812. patruzeci
813. cincizeci	814. şaizeci	815. şaptezeci	816. optzeci
817. nouăzeci	818. sută	819. mie	820. a completa
821. corect	822. a decide	823. adevărat	824. fals
825. a potrivi	826. dialog	827. a ordona	828. a încercui
829. substantiv	830. verb	831. adjectiv	832. articol
833. prepoziție	834. pronume	835. roșcat	836. cărunt
837. a oferi	838. ție	839. nouă	840. vouă
841. am	842. ai	843. a	844. ați
845. au			

## **CONCLUZII**

După cum am evidențiat în partea teoretică a lucrării noastre, necesitatea unui inventar lexical al limbii române ca limbă străină/limbă a doua este subliniată în mai multe rânduri, în ultima perioadă, de către diferiți specialiști români. Întrucât materialele care ar putea acoperi parțial această temă sunt, în principal, realizate pe baza experienței didactice a autorilor (multe dintre acestea fiind chiar inactuale și neconforme cu CECRL), ne-am propus elaborarea inventarului lexical al limbii române ca LS/L2, pentru nivelul A1, pornind de la corpusul vorbitorilor nonnativi.

Pentru aceasta, am avut nevoie de un corpus ca să ne asigurăm de validitatea elementelor incluse. Astfel, am elaborat corpusul de producții scrise ale vorbitorilor nonnativi de limba română ca limbă străină/ca limbă a doua, nivel A1, pe care l-am analizat, împreună cu cel oral (CORLS), într-o primă etapă, statistic, cu ajutorul calculatorului, în scopul obținerii coeficienților aferenți frecvenței absolute (FA) și răspândirii (R), pentru fiecare element lexical diferit. Am continuat analiza cu o etapă în care am efectuat și o analiză manuală, pentru verificare și validare, având în vedere că, la acest nivel, nu putem vorbi despre corectitudine în scriere, ceea ce a presupus o multitudine de decizii privind includerea anumitor forme în inventar.

În analiza noastră, ne-am concentrat pe ceea ce poate face și pe ceea ce poate înțelege vorbitorul nonnativ, așadar, pe ceea ce ar trebui să conțină microlimba la care acesta e expus, din punct de vedere lexical, la nivelul A1. În acest scop, pe lângă criteriile statistice, am folosit și unele cu caracter subiectiv – criteriul necesității, al întregirii sferelor lexicale sau al subiectivității. Am încercat să reducem implicarea acestora pentru a asigura un caracter empiric datelor noastre. Totuși, au fost unele situații în care criteriile statistice nu au fost suficiente, iar intervenția factorului subiectiv a fost necesară. În final, inventarul lexical al limbii române ca LS/L2, nivel A1, conține un număr de 845 de cuvinte.

În urma analizei noastre, am constat următoarele:

În ceea ce privește comparația cu situația altor limbi, am observat că pot fi identificate unele asemănări. În primul rând, dimensiunea inventarului este similară cu cea a altor liste internaționale de cuvinte (de exemplu, lista de cuvinte esențiale a lui Dang și Webb este formată din 800 de cuvinte). Deși, în proiecția făcută inițial în legătură cu dimensiunea inventarului (vezi capitolul 5), am presupus că numărul de elemente se va situa în intervalul 920-1428 de cuvinte, nu considerăm că, în forma finală, numărul elementelor lexicale ar fi prea mic. Oricum, dacă se calculează și

listele suplimentare, numărul elementelor va crește simțitor. În al doilea rând, aplicând legea lui Zipf, ni s-a confirmat ceea ce se arată în cazul altor limbi, și anume că doar primele câteva cuvinte au o frecvență ridicată, după care FA scade semnificativ, însă într-un mod predictibil. O altă asemănare este cea privind elementele conținute în inventar. Am observat că, și în cazul limbii române, primele cuvinte, ca frecvență, sunt, în principal, cuvintele de legătură. De exemplu, printre primele 10 cuvinte din inventarul lexical al limbii române ca LS/L2, se află 4 prepoziții (la,  $\hat{n}$ , de, cu), o conjuncție (si), articolul nedefinit pentru feminin, singular (o), pronumele personal, persoana I, singular (eu), adjectivul pronominal posesiv (meu, mea) și verbele a fi și a avea.

Am constatat că rangul unor cuvinte a fost influențat de temele de la proba de scriere/vorbire. Amintim aici doar cuvintele cu rang mare, precum: *a-i plăcea*, *cameră*, *da* sau *ok*. Totuși, întrucât nu am avut la dispoziție o altă metodă de a stabili care ar fi rangul potrivit pentru acestea, am decis să nu modificăm în niciun fel situația în inventar, deoarece acestea au, într-adevăr, o mare importanță în vocabularul de nivel A1.

Au întâlnit situații în care unele cuvinte au fost frecvent folosite, însă, de cele mai multe ori, într-o formă greșită, dar și într-un mod inadecvat contextului. Este cazul substantivelor *drag* și *dragă*, al adjectivului *drag* și al locțiunii adverbiale *cu drag*, care, din punctul nostru de vedere, nu caracterizează nivelul A1 și, în consecință, nu au fost incluse în inventar. O situație similară este și în cazul locuțiunii conjuncționale *după ce*, pe care, la fel, nu am inclus-o în inventar, deși, din experiența noastră didactică, știm că este predată la nivelul A1. De fapt, în acest caz, se confirmă progresia gramaticală propusă prin *Descrierea minimală*, unde locuțiunea *după ce* este introdusă abia la nivelul A2.

Întrucât am constatat că expresiile stării, așa cum le sunt prezentate studenților străini verbele cu pronume în dativ, sunt mult prea puțin utilizate, credem că nu toate caracterizează nivelul A1. De aceea am propus spre a fi incluse în inventarul lexical doar următoarele verbe: *a-i fi foame*, *a-i fi sete*, *a-i fi somn*.

Având în vedere că, la acest nivel, bagajul lexico-gramatical al unui vorbitor nonnativ este destul de redus, considerăm că substantivele proprii pot reprezenta o problemă, putând împiedica înțelegerea. Din acest motiv și urmând exemplul altor limbi, am decis formarea unei liste suplimentare, care să conțină principalele tipuri de substantive proprii întâlnite la acest nivel. Astfel, am inclus în această listă continente, țări, orașe, nume de persoane, personalități și locuri din Cluj. După cum se poate vedea din categoriile incluse, în această formă, lista este reprezentativă pentru vorbitorii nonnativi care studiază în Cluj-Napoca. Acesta este, de fapt, unul

dintre motivele pentru care am decis să nu includem elementele din această listă în inventarul propriu-zis, ci să aibă doar un caracter orientativ, fiind adaptabilă, în funcție de centrul universitar în care este folosită. Aceeași situație considerăm că apare și în cazul acronimelor și al abrevierilor, pe care le-am inserat în aceeași listă, care poate fi suplimentată sau simplificată, de fiecare dată, în funcție de grupul-țintă căruia îi este adresată.

În urma analizei calitative pe baza corpusului de producții scrise, am identificat grupuri de cuvinte specifice nivelului A1, folosite de cele mai multe ori corect. Este vorba despre structuri de tipul: cu + persoană (cu prietenul, cu prietenii sau cu prietenii mei) sau cu + obiect, în special, mijloc de transport: cu mașina, cu autobuzul. Tot din această categorie fac parte și grupurile de substantive, conectate prin prepozițiile de și cu: sală de sport, facultatea de litere, temă de casă, carne de pui, cafea cu lapte, orez cu pui, spaghete cu pui, pâine cu unt etc. sau construcțiile de la... (până) la..., din... până in... sau de... până... Am constatat că sintagmele posesive sunt folosite, în general, corect, dacă se află în situații binecunoscute (mama mea, tatăl meu, prietenul meu).

Deși, în grila de evaluare pentru producerea scrisă din *Companion Volume*, la nivelul A1, se arată că, în scriere, se folosesc doar conectori simpli (*și*, *apoi*, *după*), corpusul ne arată că totuși există o varietate mai mare. Se folosesc, des și, în general, corect, conectori, precum: *pentru că*, *după aceea*, *în general*, *de obicei*, *în final*, *de asemenea*, *de exemplu*.

Am identificat și structuri care, din punct de vedere gramatical, nu sunt predate la nivelul A1, însă care, par a fi necesare, din rezultatele bazate pe corpus. Ne referim aici la structurile *a avea nevoie SĂ/a avea chef SĂ*. Aceeași situație poate fi întâlnită și în cazul pronumelui și al adjectivului pronominal demonstrativ, prevăzut a fi predat la nivelul A2, din punct de vedere gramatical, însă, din perspectivă lexicală, cel puțin unele forme par a fi necesare și la A1 (*asta, acest, aceasta/această*), dar și în cazul verbului *a iubi*, care apare destul de frecvent (50 de ocurențe).

În ceea ce privește structurile formate din *verb* + *prepoziție* (merg LA/ÎN, plec LA, mă duc LA, vorbesc CU/DESPRE/LA/ÎN etc.), am avansat ideea că acestea sunt memorate astfel, datorită expunerii repetate la un input și la un output ce le conține. Considerăm că aceeași situație o întâlnim și în cazul unor structuri de tipul: *la mare*, *la munte*. Tot din categoria formulelor memorate fac parte, în opinia noastră, și structuri precum: *în fiecare zi*, *ne vedem*, *pe curând*, *te pup* sau chiar enunțuri întregi: *Ce mai faci?*, *Cum e vremea?*, *Orașul e frumos*. etc.

Constatând existența unui număr mare de formule memorate în corpusul vorbitorilor nonnativi de limba română ca LS/L2, ni s-a confirmat descrierea din grila de evaluare a producerii scrise din *Companion Volume*, pentru nivelul A1, care evidențiază utilizarea "unor *propoziții model*, dintr-un *repertoriu memorat*".

După cum aminteam mai sus, pe lângă criteriile statistice, am utilizat și unele cu caracter subiectiv, cum e cazul criteriului întregirii sferelor lexicale. Pentru acest criteriu, am luat în considerare 14 sfere lexicale: culori, părți ale corpului, timp, momente ale zilei, anotimpuri, zilele săptămânii, lunile anului, numerele, grade de rudenie, animale, politețe, sporturi, mijloace de transport și obiecte din clasă. Spre deosebire de lista substantivelor proprii, în acest caz, nu am extras cuvintele din inventar, deoarece considerăm că ele sunt specifice nivelului A1. Totuși, a fost nevoie, ca, în unele cazuri, să întregim aceste categorii, pentru a nu separa cuvintele care sunt, de obicei, predate/învățate/evaluate împreună. De exemplu, în categoria lunilor anului, din inventar au lipsit lunile martie, aprilie și mai, pe care le-am adăugat noi, pentru a întregi acest grup de cuvinte. Pe de altă parte, în cazul temei părțile corpului, în inventar am găsit doar trei elemente, ochi, păr și dinți, situație influențată, și în acest caz, de temele de la scriere. De aceea am decis întregirea sferei cu: cap, ureche, gură, nas, față, gât, mână, picior, deget.

Criteriul necesității vizează cuvintele utile pentru a înțelege sarcinile din cadrul cursurilor și al examenelor. După confruntarea listei de frecvență obținută pe baza unui mic corpus de cerințe din materialele de predare și evaluare folosite la Departamentul de limbă, cultură și civilizație românească, de la Universitatea Babeș-Bolyai din Cluj-Napoca, cu lista finală de frecvență din corpusurile nonnativilor, am constatat necesitatea suplimentării celei din urmă liste cu încă 15 cuvinte: a completa, corect, a decide, adevărat, fals, a potrivi, dialog, a ordona, a încercui, substantiv, verb, adjectiv, articol, prepoziție si pronume.

Pe lângă aceste criterii, am folosit și factorul subiectiv, prin care am inclus în listă următoarele elemente: *roșcat*, *cărunt*, *a oferi*, *ție*, *nouă*, *vouă*, *am*, *ai*, *a*, *ați*, *au*. Am procedat astfel, deoarece, la o ultimă verificare a inventarului, am constatat lipsa acestora, cu toate că majoritatea sunt necesare prin prisma categoriilor gramaticale specifice nivelului A1.

Prin urmare, considerăm că, prin cercetarea noastră, îmbogățim domeniul RLS, iar inventarul lexical al limbii române ca limbă străină/limbă a doua, nivel A1, și corpusul de producții scrise ale vorbitorilor nonnativi de limba română, pentru nivelul A1, reprezintă două instrumente lingvistice utile în stabilirea unui grad ridicat de unitate în ceea ce presupune înțelegerea și definirea profilului interlimbii și al

microlimbii de nivel A1 pentru limba română, indiferent de centrul de predare sau de evaluare în care acest material este utilizat.

Cu toate acestea, considerăm că, prin cercetări viitoare, se poate extinde corpusul, atât ca dimensiune, cât și ca diversitate a temelor sau a centrelor universitare din care se colectează datele. O astfel de etapă va conduce la o nouă fază, de validare a inventarului sau chiar la o ajustare a acestuia, în funcție de noile date. Necesară este extinderea sa și pe verticală, pentru fiecare nivel, situație care va permite efectuarea unor analize longitudinale în scopul evidențierii modului în care se dezvoltă limba unui vorbitor nonnativ de la un nivel la altul.

## **BIBLIOGRAFIE**

ALDERSON, C. J., "The CEFR and the need for more research", in *Modern Language Journal*, no. 91, 2007, pp. 660-662.

ALDERSON, J. C., Assessing Reading, Cambridge University Press, 2000.

ANDERSON, R. C., FREEBODY, P., *Vocabulary Knowledge*, în John T. Guthrie (Ed.), *Comprehension and Teaching: Research Reviews*, Newark, Delaware, International Reading Association, 1981, pp. 71-117.

ANDREI, M., GHIŢĂ, I., *Limba română. Fonetică, lexicologie, morfosintaxă*, București, Editura Corint, 1996.

ANTHONY, L., *AntWordProfiler* (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University, 2014, https://www.laurenceanthony.net/software.

ARIEȘAN, A., "Inventarul lexical al limbii române ca limbă străină – Propunere de cercetare", în Elena Platon, Antonela Arieșan (eds.), *Actele conferinței aniversare "40 de ani de limba română ca limbă străină la UBB. 1974-2014*", Cluj-Napoca, Editura Casa Cărții de Știință, 2014, pp. 248-250.

ARIEŞAN, A., "Observații asupra utilității inventarelor lexicale", în vol. Înspre și Dinspre Cluj: Contribuții lingvistice. Omagiu profesorului G. G. Neamțu la 70 de ani, coord. Ionuț Pomian, ed. Nicolae Mocanu, Cluj-Napoca, Editura Scriptor și Editura Argonaut, 2015, pp. 54-61.

ARIEŞAN, A., VASIU, L.I., "On the Necessity of a Learner Corpus – Romanian as a Foreign Language (RFL)", în *Globalization and National Identity. Studies on the Strategies of Intercultural Dialogue*, ed. Iulian Boldea, Arhipelag XXI, Târgu Mureş, 2016, pp. 955-966. ATKINSON, D., *Teaching in the target language: a problem in the current orthodoxy*, în "The Language Learning Journal", 8:1, 1993, pp. 2-5, DOI: 10.1080/09571739385200261.

AUSTIN, J. L., *How to do things with words*, London, Oxford University Press, 1962.

BAUER, L., NATION, I. S. P., "Word families", în *International Journal of Lexicography*, 6(4), 1993, pp. 253–279.

BIBER, D., "Representativeness in corpus design", in *Literary and Linguistic Computing*, 8(4), 1993, pp. 243–257.

BIDU-VRĂNCEANU, A. et alli, *Dicționar de Științe ale limbii (DȘL)*, București, Nemira, 2005.

BIDU-VRÂNCEANU, A., Narcisa Forăscu, *Cuvinte și sensuri*, București, Editura Științifică și Enciclopedică, 1988.

BIDU-VRÂNCEANU, A., Structura vocabularului limbii române contemporane. Probleme teoretice și aplicații practice, București, Editura Științifică și Enciclopedică, 1986.

BIRIŞ, G., BURLACU, D. V., ŞOŞA, E., Antonime. Sinonime. Analogii. Vocabular minimal al limbii române (cu traducere în limba engleză), București, Saeculum I.O., 2011.

BIRIŞ, G., Vocabularul minimal al limbii române: stadiul cunoștințelor, perspective, în Elena Platon, Antonela Arieșan (eds.), Noi perspective în abordarea românei ca limbă străină/ca limbă nematernă, Cluj Napoca, Editura Casa Cărții de Știință, 2012, pp. 173-181.

BOGAARDS, Paul, Le vocabulaire dans l'apprentissage des langues étrangères, Paris, Didier, 1994.

BRÂNCUŞ G. (coord.), IONESCU, A., SARAMANDRU, M., *Limba română: Manual pentru studenții străini* (ed. a 5-a), București, Editura Universității din București, 2003.

BREZINA, V., GABLASOVA, D., "Is there a core general vocabulary? Introducing the New General Service List.", în *Applied Linguistics*, 36(1), 2015, pp. 1–22.

BREZINA, V., Statistics in Corpus Linguistics. A practical Guide, Cambridge University Press, 2018.

BROWN, D., "An improper assumption? The treatment of proper nouns in text coverage counts", în *Reading in a Foreign Language*, 22, 2010, pp. 355–361.

BRYSBAERT, M., NEW, B., "Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. Behavior", în *Research Methods*, 41(4), 2009, pp. 977–990. BURLACU, D., "Repartizarea lexicului RLNM pe nivelurile CECRL – cu aplicație la ciclul primar", în Actele Conferinței de lansare a proiectului "Perfecționarea cadrelor didactice din

învățământul preuniversitar care predau limba română minorităților naționale", Cluj-Napoca, Editura Casa Cărții de Știință, 2011, pp. 231-241.

CAPEL, A., "Insights and issues arising from the English Profile Wordlists project", în *Research Notes*, Issue 41, University of Cambridge, ESOL Examinations, August 2010.

CARROLL, J. B., DAVIES, P., RICHMAN, B., *The American Heritage Word Frequency Book.* New York, Houghton Mifflin, Boston American Heritage, 1971.

CONSILIUL EUROPEI, Diviziunea Politici Lingvistice, Strasbourg, *Cadrul European Comun de referință pentru Limbi: Învățare, Predare, Evaluare*, traducere coordonată și revizuită de Gheorghe Moldovanu, Chișinău, 2003.

CONSTANTINESCU, M.-V., STOICA, G, Româna ca limbă străină. Corpus, Editura Universității București, București, 2020

(https://books.google.ro/books?id=wn0jzgEACAAJ&pg=PA22&source=gbs\_selected\_page s&cad=2#v=onepage&q&f=false).

COSTĂCHESCU, A., ILIESCU, M., Vocabularul minimal al limbii române curente. Cu indicații gramaticale complete, București, Demiurg, 1995.

COUNCIL OF EUROPE, Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors, Language Policy, 2018.

COUNCIL OF EUROPE, Modern Languages Division, Strasbourg, Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Cambridge, Cambridge University Press, 2001, https://rm.coe.int/1680459f97.

COUNCIL OF EUROPE, Reference Level Descriptions for National and Regional Languages (RLD). Guide for the production of RLD, Version 2, Language Policy Division, Strasbourg, November 2005.

COUNCIL OF EUROPE, Relating Language Examinations to the Common European Framework of Reference for Languages: Teaching, Learning and Assessing (CEFR). A Manual, Language Policy Division, Strasbourg, 2009.

COWIE, A. P., Learners' Dictionaries - Recent Advances and Developments, în Tickoo ed., 1989, p. 47.

COXHEAD, A., NATION, P., SIM, D., "Creating and trialling six versions of the Vocabulary Size Test", în *The TESOLANZ Journal*, 22, 2014, pp. 13–27.

CRISTEA, D., Resurse lingvistice și tehnologii ale limbajului natural. Cazul limbii române. Prelegerile Academiei Române. Filiala Iași, 2005a.

DAFINOIU, C.-V., PASCALE, L.-E., *Limba română*. *Manual pentru studenții străini din anul pregătitor (Nivel A1-A2)*, ediția a 3-a, București, Editura Universitară, 2015.

DALLER, H., MILTON, J., TREFFERS-DALLER, J., *Editor's Introduction: Conventions, terminology and an overview of the book*, în Helmut Daller, James Milton, Jeanine Treffers-Daller, *Modelling and Assessing Vocabulary Knowledge*, Cambridge, Cambridge University Press, 2007, pp. 1-32.

DANCIU, A., POANTĂ, I., *Exerciții lexicale. Limba Română*, Universitatea din Cluj-Napoca, Facultatea de Filologie, Colectivul de limba română pentru studenți străini, pentru uzul studenților, 1984.

DANG, T.N.Y., WEBB, S., "Making an essential word list for beginners", in I.S.P. Nation, op. cit, 2016, p. 153-169.

DASCĂLU JINGA, L. (coord.), *Româna vorbită actuală (ROVA). Corpus și studii*, Academia Română, Institutul de Lingvistică "Iorgu Iordan – Al. Rosetti", București, Editura Academiei Române, 2011.

DASCĂLU JINGA, L., *Corpus de română vorbită (CORV). Eşantioane*, București, Oscar Print, 2002.

DAYRELL, C., URRY, J., "Mediating climate politics: The surprising case of Brazil", în *European Journal of Social Theory* 18(3), 2015, pp. 257–273, https://doi.org/10.1177/1368431015579962.

DEX, Dictionarul explicativ al limbii române, București, Univers Enciclopedic, 1998.

DILLER, K. C., *The Language Teaching Controversy*. Rowley, Mass.: Newbury House, 1978.

DOROBĂŢ, A., FOTEA, M., Româna de bază. Essential Romanian (2 vol.), Iași, Institutul European, 1999.

ELLIS, R., BARKHUIZEN, G., *Analysing Learner Language*, Oxford University Press, 2005.

FLEMING, M., *The use of descriptors in teaching, learning and assessment*, Consiliul Europei, Divizia de Politici lingvistice, 2009.

GABLASOVA, D., BREZINA, V., MCENERY, T., "The Trinity Lancaster Corpus. Development, description and application", in *International Journal of Learner Corpus Research* 5:2 (2019), pp. 126–158.

GAIRNS R., REDMAN, S., Working with words. A guide to teaching and learning vocabulary, University of Cambridge, Cambridge, 1986

GALR, *Gramatica limbii române*, vol. I *Cuvântul*, vol. II *Enunţul*, Editura Academiei Române, Bucureşti, 2005.

GASS, S. M., SELINKER, L., Second Language Acquisition. An introductory Course (Third Edition), New York/London, Routledge, 2008.

GBLR, Gramatica de bază a limbii române (coord. Gabriela Pană Dindelegan), Editura Univers Enciclopedic Gold, București, 2010.

GHEORGHE, M., MĂDA, S., SĂFTOIU R., Comunicarea la locul de muncă. Corpus de interacțiune verbală în mediul profesional (+ CD), Brașov, Editura Universității Transilvania din Brașov, 2009.

GILQUIN, G., "From design to collection of learner corpora", în S. Granger, G. Gilquin, F. Meunier (eds.), *The Cambridge Handbook of Learner Corpus Research*, 2015, pp. 12-22.

GILQUIN, G., GRANGER, S., "How can data-driven learning be use in language teaching?", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, pp. 359-370.

GRAUR, Al., Fondul principal al limbii române, București, Editura Științifică, 1957.

GREEN, A., "The theoretical foundations for functions in the Council of Europe modern languages projects and the Common European Framework of Reference for languages", in Michael Milanovic, Nick Saville (eds.), *Language Functions Revisited*, Cambridge, Cambridge University Press, 2012, pp. 7-38.

GRIES, S. Th., Ten lectures on corpus Linguistics with R: applications for usage-based and psycholinguistic research, Leiden, Boston, Brill, 2020.

HALL, C. J., "The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development", in *International Review of Applied Linguistics in Language Teaching*, 2002, 40, pp. 69–87, disponibil online: http://doi.org/cntmwz.

HARRISON, J., BARKER, F. (eds.), *English Profile Studies. English Profile in Practice*, Cambridge University Press, Cambridge, 2015.

HOARŢĂ CĂRĂUŞU, L. (coord.), *Corpus de limbă română vorbită actuală*, Iași, Editura Universității "Alexandru Ioan Cuza", 2013.

HOARŢĂ CĂRĂUŞU, L. (coord.), *Corpus de limbă română vorbită actuală*, Iași, Editura Tehnică, Științifică și Didactică Cermi, 2005.

HRISTEA, Th., *Introducere în studiul vocabularului*, în Theodor Hristea (coord.), *Sinteze de limba română*, București, Editura Albatros, 1984.

HULSTIJN, J. H., "The common European framework of reference for languages. A challenge for applied linguistics", in *International Journal of Applied Linguistics*, 165, 2013, pp. 3–18.

HULSTIJN, J. H., "The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency", in *Modern Language Journal*, 9, 2007, pp. 663–667.

HUNT, A., BEGLAR, D., "Current Research and Practice in Teaching Vocabulary", în Richards, Jack C., Renandya, Willy A., *Methodology in Language Teaching. An Anthology of Current Practice*, Cambridge University Press, Cambridge, 2002, pp. 258-264.

ILIESCU, M. (coord.), Vocabularul minimal al limbii române pentru studenți străini, București, Editura Didactică și Pedagogică, 1981.

ION, R., IRIMIA, E., ŞTEFĂNESCU, D., TUFIŞ, D., *ROMBAC: The Romanian Balanced Annotated Corpus*, în *LREC*, 2012, pp. 339-344.

IONESCU-RUXĂNDOIU, L. (coord.), *Interacțiunea verbală (IV II). Aspecte teoretice și aplicative. Corpus*, București, Editura Universității din București, 2007.

IONESCU-RUXĂNDOIU, L. (coord.), *Interacțiunea verbală în limba română actuală*. *Corpus (selectiv)*. *Schiță de tipologie*, București, Editura Universității din București, 2002.

IONILĂ, F.-L., AVRAM, P., Dicționar minimal al limbii române cu indicații gramaticale uzuale și traducere în franceză, engleză, spaniolă, București, Niculescu, 1996.

IRIMIA, E., "Accelerarea dezvoltării unui corpus digital adnotat cu relații de dependență pentru limba română utilizând resurse și instrumente construite pentru alte limbi", în *Revista Română de Informatică* și *Automatică*, vol. 25, nr. 3, 2015, pp. 5 -16.

JUILLARD, A., EDWARDS, P.M.H., JULLIARD, I., *Frequency dictionary of Romanian Words*, Mouton & Co, The Hague, 1965.

KIEFER, P. (coord.), Goethe-Zertifikat A1, Start Deutsch 1, Prüfungsziele Testbeschreibung, Goethe Institut, 2011.

KOBELEVA, P. P., "Second language listening and unfamiliar proper names: Comprehension barrier?", în *RELC Journal*, 43(1), 2012, pp. 83–98.

KOHN, D., Puls. Manual de limba română pentru străini. Curs / Exerciții, Iași, Polirom, 2009. KOIZUMI, R., "Vocabulary and Speaking", în *The Encyclopedia of Applied Linguistics*, Oxford, John Wiley and Sons, Inc., 2013, pp. 1–7.

KUSSELING, F., DECOO, W., Europe language learning: The challenges of comparable assessment, în European Studies Conference Selected Proceedings, University of Nebraska, Omaha, 2009.

LADO, R., *Language Teaching. A scientific Approach*, New York, McGraw-Hill, Inc, 1964. LAUFER, B., RAVENHORST-KALOVSKI, G. K., "Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension", în *Reading in a Foreign Language*, April 2010, Volume 22, No. 1, pp.15-30.

LEECH, G., *New resources, or just better old ones? The Holy Grail of representativeness* in Marianne Hundt, Nadja Husselhauf, Carolin Biewer (eds.), *Corpus Linguistics and the Web*, Amsterdam, Rodopi, 2007, pp. 133-150. https://doi.org/10.1163/9789401203791\_009.

LITTLE, D., "Common European Framework of Reference for languages: content, purpose, origin, reception and impact", în *Language Teaching*, no. 39, 2006, pp. 167-190.

LITTLE, D., ERICKSON, G., "Learner identity, learner agency, and the assessment of language proficiency: Some reflections prompted by the Common European framework of

reference for languages", in *Annual Review of Applied Linguistics*, no. 35, 2015, pp. 120-139.

LITTLE, D., The Common European Framework of Reference for Languages, the European Language Portfolio, and language learning in higher education, pp 1-19, http://languagesinitiative.ie/images/CEFR\_ELP.pdf.

LOWIE, W. M., "The CEFR and the dynamics of second language learning: Trends and challenges", în Language Learning in Higher Education. Journal of the European Confederation of Language Centres in Higher Education (CercleS), 2(1), 2012, Berlin/Boston, De Gruyter, pp. 17–34.

MACREA, D., *Probleme ale structurii și evoluției limbii române*, București, Editura Științifică și Enciclopedică, 1982.

MARELLO, C., "Word lists in Reference Level Descriptions of CEFR (Common European Framework of Reference for Languages)", în *Proceedings of the XV Euralex International Congress*, Department of Linguistics and Scandinavian Studies University of Oslo, 2012, pp. 328-335.

MCCARTHY, M., O'KEEFFE, A., Historical perspective. What are corpora and how have they evolved in Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012.

MCENERY, T., Course in Corpus Linguistics: Method, Analysis, Interpretation (curs online), Lancaster University.

MCENERY, T., HARDIE, A., *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge, 2012.

MCENERY, T., HARDIE, A., *Corpus linguistics: method, theory and practice*, Cambridge University Press, 2011.

MILTON, J., ALEXIOU, T., Developing a vocabulary size test in Greek as a foreign language (prezentare ppt).

MILTON, J., The development of vocabulary breath across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidlines, examinations, and textbooks across Europe, in Eurosla Monographs series 1, Communicative proficiency and linguistic development, 2010, pp. 211-232.

MILTON, J., *Measuring Second Language Vocabulary Acquisition*, Bristol, Buffalo, Toronto, Multilingual Matters, 2009.

MÎRZEA VASILE, C., "Corpusurile lingvistice de limba română și importanța lor în realizarea de materiale didactice pentru limba română ca limbă străină" (lucrare prezentată la Conferința aniversară 40 de ani de limba română ca limbă străină la UBB.1974-2014, Facultatea de Litere, UBB, Cluj-Napoca, 3-4 octombrie 2014).

MOLDOVAN, V., POP, L., URICARU, L., *Nivel prag pentru învățarea limbii române ca limbă străină*, Strasbourg, Consiliul de Cooperări Culturale, 2001.

MÜLLER-HARTMANN, A., SCHOCKER-VON DITFURTH, M., *Introduction to English language Teaching*, Stuttgart, Klett Lerntraining, 2010.

MUREȘANU, I., "Un dicționar practic pentru studenți străini – abordare statistică", în *Lexicologie și lexicografie didactică*, Timișoara, Tipografia Universității din Timișoara, 19985, p. 24-28.

MUREȘANU, I., Polisemia în limba română, Cluj-Napoca, Casa Cărții de Știință, 2005.

NAGY, E., ANDERSON, R. C., "How many words are there in printed school English?", în *Reading Research Quarterly*, 19(3), 1984, pp. 304–330.

NATION P., WARING R., "Vocabulary size, text coverage and word lists", în Schmitt, N., McCarthy, M. (eds.), *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge, Cambridge University Press, 1997, pp. 6-19.

NATION, I. S. P., *Learning Vocabulary in Another Language*. (Second edition), Cambridge, Cambridge University Press, 2013.

NATION, I.S.P., DEWEERDT, J., "A defence of simplification", in *Prospect* 16(3), 2001, pp. 55-67.

NATION, I.S.P., *Learning vocabulary in another language*, Cambridge, Cambridge University Press, 2001.

NATION, I.S.P., *Making and Using Word Lists for Language Learning and Testing*, John Benjamin Publishing Company, Amsterdam/Philadephia, 2016.

NATION, P., BEGLAR, D., "A vocabulary size test", în *The Language Teacher*, 31(7), 2007, pp. 9–13.

NATION, P., CRABBE, D., "A survival language learning syllabus for foreign travel", în *System*, 19(3), 1991, pp. 191–201.

NATION, P., KOBELEVA, P., "Proper nouns", în I.S.P. Nation, *Making and Using Word Lists for Language Learning and Testing*, John Benjamin Publishing Company, Amsterdam/Philadephia, 2016, pp. 55-64.

NEAMȚU, G. G., *Studii și articole gramaticale*, Editura Napoca Nova, Cluj-Napoca, 2014. NEAMȚU, G. G., Teoria și practica analizei gramaticale. Disticții și... distincții, Editura Excelsior, Cluj-Napoca, 1999.

NELSON, M. "Building a written corpus. What are the basics?", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, pp. 53-65.

NORTH, B., "Putting the Common European Framework of Reference to good use", in *Language Teaching*, 47(02), 2014, pp. 228–249.

NORTH, B., "The CEFR illustrative descriptor scales", in *Modern Language Journal*, no. 91, 2007, pp. 656-659.

NORTH, B., "The Relevance of the CEFR to Teacher Training", in *Babylonia*, no. 2, 2008, Zürich, pp. 55-57, http://clients.squareeye.net/uploads/eaquals/Baby2-08CECRL-North.pdf. North, B., Inventaire linguistique des contenus clés aux niveaux du CECRL, Eaquals, 2015. NORTH, B., ORTEGA, A., SHEEHAN, S., *A Core Inventory for General English*, Eaquals, 2014.

NORTH, B., Perspectives on Language Proficiency and Aspects of Competence: a reference paper defining categories and levels, Strasbourg, Council of Europe CC-Lang (94)20, 1997.

O'KEEFEE, A., MCCARTHY, M., *From Corpus to Classroom*, Ronald Carter, Cambride University Press, Cambridge, 2007.

OGDEN, C. K., *The Basic Words*, London, Kegan Paul, Trench, Trubner & Co., 1932, http://ogden.basic-english.org/texts/tbw.html.

PĂLTINEANU, V., Dicționar practic pentru studenți străini. 1000 de cuvinte românești, Universitatea Babeș-Bolyai, Cluj-Napoca, 1980.

PICCARDO, E. et alli, *Pathways Through Assessment, Learning and Teaching in the CEFR*, Graz, Austria, European Centre for Modern Languages, 2011.

PLATON, E, "The idea of progression in designing the curriculum of romanian as a foreign language (RFL)", în "Studia UBB Philologia", LXIV, 2, 2019a, pp. 141-156, DOI:10.24193/subbphilo.2019.2.12.

PLATON, E. (coord.), *Româna ca limbă străină (RLS). Caiete didactice. A1*+, Cluj-Napoca, Editura Casa Cărtii de Stiintă, 2012.

PLATON, E., (coord.), ARIEȘAN, A., FODOR, L., *Evaluarea limbii române la ciclul primar: EVRO P5*, Cluj-Napoca, Editura Casa Cărții de Știință, 2015.

PLATON, E., "Descrierea minimală a limbii române (A1, A2, B1, B2) – un instrument lingvistic necesar în domeniul RLS/RLM, în Elena Platon, Ioana Sonea, Lavinia Vasiu, Dina Vîlcu, *Descrierea minimală a limbii române. A1, A2, B1, B2*, Cluj-Napoca, Editura Casa Cărții de Știință, 2014, pp. 3-10.

PLATON, E., "Reflexii asupra conceptului de interlimbă", în vol. *Înspre și Dinspre Cluj: Contribuții lingvistice. Omagiu profesorului G. G. Neamțu la 70 de ani*, coord. Ionuț Pomian, ed. Nicolae Mocanu, Cluj-Napoca, Editura Scriptor și Editura Argonaut, 2015, pp. 527-539. PLATON, E., "RLS - Validarea competențelor lingvistice în conformitate cu Cadrul European Comun de Referință și cu Pașaportul lingvistic European", in vol. *România Orientale*, no. XXI, Roma, Bagatto Libri, 2008, p. 306.

PLATON, E., "The «micro-language» – an assault on natural languages?", în *Lingua*. *Language and culture*, XVIII, nr. 2, 2019b, pp. 67-83.

PLATON, E., GOGÂȚĂ, C., VASIU, L.-I., URSA, A., România în 50 de povești. Limba română ca limbă străină (RLS). Texte gradate A1-A2, Editura Presa Universitară Clujeană, Cluj-Napoca, 2021.

PLATON, E., Repere pentru delimitarea nivelurilor de competență lingvistică în RLS, în G.G. Neamțu, Ștefan Gencărău, Adrian Chircu, Limba Română – Abordări tradiționale și moderne, Cluj-Napoca, Presa Universitară Clujeană, 2009, pp. 499-505.

PLATON, E., Româna ca limbă străină (RLS). Elemente de metadidactică, Cluj-Napoca, Editura Presa Universitară Clujeană, 2021.

PLATON, E., SONEA, I., VASIU, L., VÎLCU, D., *Descrierea minimală a limbii române.* A1, A2, B1, B2, Cluj-Napoca, Editura Casa Cărții de Știință, 2014.

PLATON, E., SONEA, I., VÎLCU, D., Exerciții audio: Româna ca limbă străină (RLS): A1, A2, B1, B2, C1, C2, Cluj-Napoca, Editura Efes, 2010.

PLATON, E., SONEA, I., VÎLCU, D., Manual de limba română ca limbă străină (RLS). A1-A2, Cluj-Napoca, Editura Casa Cărții de Știință, 2012.

PLATON, E., *Un proiect utopic de lexicografie didactică în RLS*, în Elena Platon, Cristina Bocoş, Diana Roman, Lavinia Vasiu (editori), *Discurs polifonic în româna ca limbă străină (RLS)*, nr. 2/2020, Editura Presa Universitară Clujeană, 2020, pp. 357-378.

POP L. (coord.), Rodica Frențiu, *Seria Autodidact. Româna: A1-B1. Caietul Studentului*, Cluj, Editura Echinox, 2008.

POP L. (coord.), Seria Autodidact. Româna: A1-B1, Cluj, Editura Echinox, 2008.

POP, L. (ed.), La langue virtuelle. Recherche sur les forums des jeunes, Cluj, Ed. Echinox, 'Atelier', 2008.

POP, L. (éd.), *Où va la communication? (Comunicarea, încotro?*), Cluj, Ed. Echinox, 'Atelier', 2010.

POP, L. (ed.), Verba volant. Recherches sur l'oral, Cluj, Ed. Echinox, 'Atelier', 2004.

POP, L., DUMA, M., PAŞCALĂU, C. (eds.), *Façons de parler.ro*, Cluj, Ed. Echinox, 'Atelier', 2011.

POP, L., MOLDOVAN, V., MUREŞANU, I., Româna în 15 lecții. Curs intensiv de limba română, Cluj-Napoca, Universitatea din Cluj-Napoca, 1989.

POP, L., Româna cu sau fără profesor. Le Roumain ou sans professeur. Romanian with or without a teacher, Cluj-Napoca, Echinox, 2003.

READ, J., "Plumbing the depths: How should the construct of vocabulary knowledge be defined?", în Bogaards, Paul, Batia, Leufer (eds.), *Vocabulary in a Second Language*. *Selection, Aquisition and Testing*, Amsterdam, John Benjamins Publishing Company, 2004, pp. 209-224.

READ, J., Assessing Vocabulary, Cambridge, Cambridge University Press, 2000.

READ, J., "Research in Teaching Vocabulary", în *Annual Review of Applied Linguistics*, 24, 2004, pp. 146-161.

REPPEN, R., "Building a corpus. What are the key considerations?", în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, pp. 31-37.

SACALÎŞ, L., *Un model de lucru în DERLS*, la Conferința internațională *Discurs polifonic în româna ca limbă străină (RLS)*, 29 octombrie 2021, Universitatea Babeș-Bolyai, Cluj-Napoca (comunicare).

SALA, M. (coord.), *Vocabularul reprezentativ al limbilor romanice*, Editura Științifică și Enciclopedică, București, 1988.

SCHMITT, N., SCHMITT, D., "A reassessment of frequency and vocabulary size in L2 vocabulary Teaching", în *Language Teaching*, 47(4), 2014, pp. 484–503.

SCHMITT, N., SCHMITT, D., *Vocabulary in language teaching. Second Edition*, Cambridge, Cambridge University Press, 2020.

SEARLE, J. R. Speech Acts. An Essay in the Philosophy of Language, Cambridge, Cambridge University Press, 1969.

SEMINO, E. et. alli, "The online use of violence and journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study", în *BMJ Supportive & Palliative Care* 7(1), 2017, pp. 60–66, https://doi.org/10.1136/bmjspcare-2014-000785.

SEVIER, M., "The Compleat Tutor, v.4", in *Teaching English as a Second or Foreign Language*, vol. 8, no. 3, december 2004.

SINCLAIR, J. M. (Ed), *Collins COBUILD English Language Dictionary*. London: Collins, 1987.

SINCLAIR, J. M., Looking Up, London, Collins ELT, 1987.

SINCLAIR, J., Trust the Text: Language, Corpus and Discourse, London, Routledge, 2004. SLAMA-CAZACU, T., *Psiholingvistica. O știință a comunicării*, București, Editura All, 1999.

SONEA, I., "The impact of the CEFR on Teaching and Assessing Romanian as a Second/Foreign Language. Practice and Research", in Lucia-Mihaela Grosu-Răduleascu (ed.), Foreign Language Teaching in Romanian Higher Education. Teaching Methods, Learning Outcomes, Cham, Springer, col. "Multilingual Education, no. 27, 2018, pp. 111-124.

STÆHR, L. S., "Vocabulary size and the skills of listening, reading and writing", în *Language Learning Journal*, 36(2), 2008, pp. 139-152.

STOICA, G., "Dificultăți lexico-semantice în achiziția limbii române ca limbă străină", în Elena Platon, Antonela Arieșan (eds.), *Noi perspective în abordarea românei ca limbă străină/ca limbă nematernă*, Cluj Napoca, Editura Casa Cărții de Știință, 2012, pp. 182-191. STUBBS, M., *Language corpora*, in: Davies, Alan, Catherine Elder, *The Handbook of Applied Linguistics*, Blackwell Publishing Ltd., 2013.

SUCIU, A.C., "Strategii de predare a competenței lingvistice (lexic, fonologie, morfologie și sintaxă) – Gramatică și vocabular", în Elena Platon, Diana-Viorela Burlacu, Ioana Sonea (coord.), *Procesul de predare-învățare a limbii române ca limbă nematernă (RLNM) la ciclul primar. RLNM: P1-ciclul primar*, Cluj-Napoca, Editura Casa Cărții de Știință, 2011, pp. 81-98.

SZABO, Cz., "Exploring the Mental Lexicon of the Multilingual: Vocabulary Size, Cognate Recognition and Lexical Access in the L1, L2 and L3", in *Eurasian Journal of Applied Linguistics* 2 (2) (2016), pp. 1-25.

SZABO, Cz., "Introducing a Romanian Frequency List and the Romanian Vocabulary Levels Test, in *Current Issues in Linguistic Variation: The 14th international conference of the Department of Linguistics*, Vol. 2, University of Bucharest: Bucharest University Press, 2015, pp. 303–317.

SZUDARSKI, P., Corpus Linguistics for Vocabulary. A Guide for Research, Routledge, London and New York, 2018

ȘERBAN, V., EVSEEV, I., *Vocabularul românesc contemporan*, Timișoara, Editura Facla, 1978.

TODEA, A. "Developing an Analysis Grid for B1/B2 Reading Tasks and Items", în *Studia*. *Philologia*, LXII, 2/2017, vol. 62, pp. 167-182.

TOGNINI BONELLI, E., *Theretical overview of the evolution of corpus linguistics*, în Michael McCarthy, Anne O'Keeffe (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, New York, 2012, pp. 14-27.

TRIM, J.L.M. "Prefață la Nivel Prag", în Victoria Moldovan, Liana Pop, Lucia Uricaru, *Nivel prag pentru învățarea limbii române ca limbă străină*, Strasbourg, Consiliul de Cooperări Culturale, 2001, p. 8.

TUFIȘ, D., CRISTEA, D., *An Insight into CoRoLa*, prezentare invitată la SpeD 2017 (https://profs.info.uaic.ro/~dcristea/Talks/SPeD2017%20talk\_DT\_DC.ppt.pdf).

TUFIŞ, D., IRIMIA, E., "RoCo-News: A Hand Validated Journalistic Corpus of Romanian", în Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06), 2006, pp. 869-872.

UNIVERSITY OF CAMBRIDGE, *Cambridge ESOL: Research notes*, Issue 41, August 2010, Cambridge University Press, p. 3.

UNIVERSITY OF CAMBRIDGE, Introductory Guide to the Common European Framework of Reference (CEFR) for English Language Teachers, Cambridge, Cambridge University Press, 2013, pp. 1-12.

UNIVERSITY OF CAMBRIDGE, Using the CEFR Principles of Good Practice, October 2011, Cambridge ESOL, Cambridge.

VALENTIN, N, *Variabile statistice* (*clasificare și indicatori*), Universitatea București, prezentare ppt, disponibilă pe: https://www.academia.edu/7577374/Variabile\_statistice.

VAN EK, J. A., The Threshold Level, Council of Europe, 1975.

VAN EK, J. A., TRIM, J.L.M., *Threshold 1990*, Cambridge, Cambridge University Press, 1991

VAN ZEELAND, H., Vocabulary and Listening. The Encyclopedia of Applied Linguistics, 2013, pp.1–6.

VASIU, L. I., *Achiziția limbii române ca L2. Interlimba la nivelul A1*, Cluj-Napoca, Editura Presa Universitară Clujeană, 2020.

VASIU, L. I., ARIEŞAN A., "RFL – Cohesion. Linking words and phrases (Level B2)", în I. Boldea (ed.), *Globalization and National Identity. Studies on the Strategies of Intercultural Dialogue*, Arhipelag XXI, Târgu Mureş, 2016, pp. 1074-1088.

VÎLCU, D., ARIEȘAN, A., "Proposal for illustration of the supplement with a test for end of LSP course (medical profile), la atelierul *Limbajul pentru scopuri specifice - Language for Specific Purposes (LSP)"*, (lucrare prezentată în cadrul întâlnirii bianuale ALTE Multilingvismul și utilizarea *limbajului pentru scopuri specific –* Rolul evaluării, 11-13 aprilie 2017, Cluj-Napoca).

VÎLCU, D., ARIEŞAN, A., VASIU, L.-I., "The use of test taker productions in redesigning writing assessment grids. A corpus based study", în *ALTE* (2017). *Learning and Assessment:* 

Making the Connections - Proceedings of the ALTE 6th International Conference, 3-5 May 2017, pp. 112-117.

VÎLCU, D., ARIEŞAN, A., VASIU, L.-I., "The use of test takers productions in redesigning writing assessment grids: a corpus based study" (lucrare prezentată la ALTE 6th International Conference, 3-5 mai 2017, Bologna, Italia).

WARING, R., "A study of receptive and productive learning from word cards", in *Studies in Foreign Languages and Literature* (Notre Dame Seishin University, Okayama), 21(1), 1997, pp. 94–114.

WARING, R., NATION, P., "Second language reading and incidental vocabulary learning", în *Angles on the English- speaking world*, vol. 4, 2004, pp. 97-110.

WEBB, S., "Receptive and productive vocabulary size", in *Studies in Second Language Acquisition*, 30(1), 2008, pp. 79–95.

WEBB, S., NATION, P., *How Vocabulary is Learned*, London, Oxford University Press, 2017.

WEIR, J. C., HUIZHONG, Y., YAN, J., Studies in Language Testing 12: An empirical investigation of the componentiality of L2 Reading in english for academic purposes, Cambridge University Press, 2000.

WILKINS, D.A., Linguistics in Language Teaching, London, Annold, 1972.

WISNIEWSKI, K., "Empirical learner language and the levels of the Common European Framework of Reference", in *Language Learning*, 67(S1), 2017, pp. 232–253.

YANAGISAWA, A., WEBB, S., "Involvement Load Hypothesis Plus: Creating an Improved Predictive Model of Incidental Vocabulary Learning". *Studies in Second Language Acquisition*, 2021, pp. 1-30. doi:10.1017/S0272263121000577.

ZDRENGHEA, S., *Limba română*. *Nivel A (începători)*, Cluj-Napoca, Universitatea Babeș-Bolyai, 1995.

ZENO, S. M., IVENS, S. H., MILLARD, R. T., DUVVURI, R., *The Educator's Word Frequency Guide*, Touchstone Applied Science Associates, 1995.

ZIMMERMAN, C. B., Word Knowledge. A Vocabulary Teacher's Handbook, Oxford University Press, 2009.

#### Referințe web

Consiliul Europei, URL: http://www.coe.int/t/dg4/linguistic/DNR\_EN.asp#P61\_9444 El Plan Curricular del Instituto Cervantes, URL:

http://cvc.cervantes.es/ensenanza/biblioteca ele/plan curricular/indice.htm.

*English as a Second Language, URL:* http://www.rong-chang.com/.

English Unlimited and English Profile Wordlists, URL: http://www.cambridge.org/.

English Vocabulary Profile, URL: http://www.englishprofile.org/.

Formula PERT, URL: http://www.brighthubpm.com/methods-strategies/15682-introduction-to-the-pert-formula-series/

Large English Vocabulary Wordlists, URL: http://www.manythings.org/vocabulary/lists/l/. Le Point du FLE, URL: http://www.lepointdufle.net/didactiquefle.htm.

Listă de cuvinte pentru nivelul A1, Goethe Institut, URL:

http://www.goethe.de/lrn/prf/pro/hdb/Pruefungsziele\_Testbeschreibung\_A1\_SD1.pdf *Listă de cuvinte pentru nivelul B2, Goethe Institut, URI:* 

 $http://www2.klett.de/sixcms/list.php?page=lehrwerk\_extra\&titelfamilie=Mittelpunkt\&extra=Mittelpunkt-Online\&modul=inhaltsammlung\&inhalt=klett71prod\_1.c.185053.de\&kapitel=185060$ 

Profile Deutsch, URL: http://www.osd.at/default.aspx?SIid=14&LAid=2.

Compleat Word Tutor: URL: http://www.lextutor.ca/vp/comp/.

*Text Inspector*: https://textinspector.com.

Range: https://www.wgtn.ac.nz/lals/resources/vocabulary-analysis-programs.

Cărți adaptate pe nivel: https://erfoundation.org/wordpress/graded-readers/

Extended Mark-Up Language (XML): https://www.w3.org/TR/REC-xml/

Corpus Encoding Standard (CES): https://www.cs.vassar.edu/CES/

*Text Encoding Initiative (TEI)*: https://tei-c.org. *CoRoLa*, disponibil online: http://corola.racai.ro.





ISBN: 978-606-37-1689-8